REVISITING THE MULTIDIMENSIONALITY AND MEASUREMENT OF HOSTILE MASCULINITY: TESTING BIFACTOR S-1 CFA AND ESEM MODELS

by

WENQI ZHENG

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

DOCTOR OF PHILOSOPHY IN PSYCHOLOGY

2026

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:

Michele R. Parkhill, Ph.D., Chair
Mark Manning, Ph.D.
Matt L. Miller, Ph.D.

ACKNOWLEDGEMENTS

I would first like to acknowledge the Oakland University Research Office and the Office of the Provost for funding this research. I am also grateful to Dr. Virgil Zeigler-Hill and Danielle Bildson in the Oakland University Department of Psychology for their administrative support. I would also like to thank faculty in the Department of Psychology for contributing to my development as a researcher through formal instruction and mentorship for professional growth. I cannot thank Drs. Mark Manning and Matt Miller enough for their willingness to serve on my dissertation committee, and Dr. Melissa McDonald for serving on my comprehensive examination committee. Their thoughtful feedback and guidance greatly strengthened both this research and my development as a scholar. I am especially grateful to Dr. Michele Parkhill, who served as my academic and research advisor throughout my Ph.D. education and chaired both my comprehensive examination and dissertation committees. Her endless encouragement, patience, and support helped me grow not only as an academic and researcher, but also as a person. She created a welcoming and supportive research environment in which I was able to flourish, and her care extended far beyond academics. Her guidance and kindness helped me navigate life in Michigan and made our lab feel like a home. I will always be grateful for the many ways she supported me throughout this journey. Finally, I thank my mother, Baoqing Zhang, for giving me the trust, freedom, and unconditional support to choose my own path and pursue a Ph.D. in psychology. I hope she knows how deeply I love her and how grateful I am for everything she has given me. I also thank my beloved dog, Lucyfur, who stayed beside me through countless late nights. Her companionship gave me the comfort and strength I needed to make it through the final year of my degree.

ABSTRACT

REVISITING THE MULTIDIMENSIONALITY AND MEASUREMENT OF HOSTILE MASCULINITY: TESTING BIFACTOR S-1 CFA AND ESEM MODELS

by

Wenqi Zheng

Advisor: Michele R. Parkhill, Ph.D.

The Confluence Model of sexual aggression identifies hostile masculinity (HM) as a central risk factor characterized by hostility toward women (HTW), adversarial sexual beliefs (ASB), and sexual dominance (SOD). Despite its theoretical importance, the dimensionality and measurement of HM remain insufficiently evaluated, with prior research relying on inconsistent operationalizations and composite scoring practices. This dissertation refined the measurement of HM in a three-study psychometric investigation. Study 1 applied integrative data analysis (IDA) to two existing datasets (total $N = 720$), using the graded response item response theory model (GRM-IRT) to evaluate item performance and differential item functioning (DIF) analyses to assess measurement equivalence across two datasets. Items showing weaker psychometric performance were removed to develop shortened measures of the three HM components: HTW, ASB, and SOD. Study 2 used the same pooled dataset to compare higher-order, bifactor, and bifactor S-1 specifications estimated using confirmatory factor analysis (CFA) and exploratory structural equation modeling (ESEM). Results indicated that the ASB-anchored bifactor S-1 CFA model demonstrated the best balance of model fit,

parsimony, and interpretability, $\chi^2(174) = 603.34$, CFI = .943, TLI = .931, RMSEA = .067, SRMR = .033, supporting a general HM factor anchored in ASB and three residual-specific factors representing SOD and two HTW dimensions. In Study 3 (N = 329), bivariate correlations and cross-validated regression analyses supported the convergent and nomological validity of the ASB-anchored model relative to an HTW resentment-anchored alternative and a conventional unidimensional HM total score. Together, the findings support the ASB-anchored bifactor S-1 CFA model as a parsimonious and theoretically interpretable representation of HM.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC Akaike Information Criterion

AIV Acceptance of Interpersonal Violence

ALCSEX Alcohol-Related Sexual Behavior

ASB Adversarial Sexual Beliefs

ASV Attitudes Supporting Violence

BIC Bayesian Information Criterion

CFA Confirmatory Factor Analysis

CFI Comparative Fit Index

CSA Childhood Sexual Abuse

DIF Differential Item Functioning

DRGSEX Drug-Related Sexual Behavior

EAP Expected A Posteriori

ECV Explained Common Variance

EFA Exploratory Factor Analysis

ESEM Exploratory Structural Equation Modeling

GED General Educational Development

GRM Graded Response Model

HM Hostile Masculinity

HTW Hostility Toward Women

IDA Integrative Data Analysis

IRB Institutional Review Board

IRMA-SF Illinois Rape Myth Acceptance Scale-Short Form

IRT Item Response Theory

IS Impersonal Sex

KMO Kaiser-Meyer-Olkin

MAE Mean Absolute Error

MLR Robust Maximum Likelihood Estimator

MSE Mean Squared Error

MTurk Amazon Mechanical Turk

PUC Percentage of Uncontaminated Correlations

RMSEA Root Mean Square Error of Approximation

RMSE Root Mean Squared Error

SES Sexual Experiences Survey

SES-SFP Sexual Experiences Survey-Short Form Perpetration

SOI-R Revised Sociosexual Orientation Inventory

SOD Sexual Dominance

SRMR Standardized Root Mean Square Residual

SRP-III Self-Report Psychopathy-III Scale

TEQ Toronto Empathy Questionnaire

TLI Tucker-Lewis Index

CHAPTER ONE

INTRODUCTION

Sexual violence is a pervasive global issue with serious health, social, and economic consequences. By definition, sexual violence encompasses any sexual act committed without consent, ranging from unwanted sexual contact to attempted and completed rape (Koss et al., 2024). Empirical research has documented that sexual violence victimization is associated with multiple adverse health outcomes, including physical injuries (Sugar et al., 2004), and sexually transmitted diseases (Allsworth et al., 2009), as well as chronic psychological distress such as depression, anxiety, PTSD (Choudhary et al., 2012; Johansson et al., 2024; Santaularia et al., 2014), and alcohol/drug abuse (Bird et al., 2019; Najdowski & Ullman, 2009; S. Ullman et al., 2013). Beyond individual health risks, sexual violence also engenders substantial social disruptions, impairing interpersonal functioning and perpetuating stigma and fear within communities (Deitz et al., 2015; Kennedy & Prock, 2018; Schwartz, 1991). These devastating personal and social impacts are further compounded by severe economic hardships. According to CDC research (Peterson et al., 2017), the estimated lifetime economic burden per rape victim is $122,461, contributing to a staggering societal cost of $3.1 trillion through direct expenses, lost productivity, and reduced quality of life.

Worldwide, the prevalence of sexual violence is astounding. World Health Organization (WHO; 2024) documented that approximately 30% of women have experienced either physical and/or sexual intimate partner violence or non-partner sexual violence in their lifetime. The rates of sexual violence vary significantly across regions, with the highest prevalence in low- and middle-income regions. In certain parts of Sub-Saharan Africa, South Asia, and Latin America, rates of intimate partner violence can reach as high as 33%. In the US, these rates tend to be lower, but still troubling, with about 20% of women having experienced sexual violence at some point in their lives (WHO, 2024).

While men can also be victims of sexual violence and women can be perpetrators, men remain disproportionately the perpetrators and women the victims. According to the National Sexual Violence Resource Center, 91% of victims of rape and sexual assault are female, and 9% are male (NSVRC, 2015). About 75 to 90 percent of sexual abuse is committed by men or male adolescents, while about 10 to 25 percent is perpetrated by women or female adolescents. This pattern, though, might be partially attributed to the under-reporting of men's sexual victimization and reflects the unequal power dynamics between genders, as well as deep-rooted cultural norms and values. A recent study on the U.S. state-level sexual violence provided further evidence for this pattern, finding that women reported substantially higher lifetime prevalence of completed or attempted rape than men and that state-level gender inequality was associated with higher prevalence of rape involving physical force among women (Kearns et al., 2020). Although public awareness and vigilance often center on stranger-perpetrated assault, empirical evidence indicates that sexual violence is predominantly perpetrated by acquaintances. The highest proportion of cases involves former or current intimate partners, followed by acquaintances such as friends, neighbors, family associates, cohabitants, colleagues, and classmates, with stranger-perpetrated assaults comprising the smallest percentage (Basile et al., 2022).

Over the past decades, psychological research in sexual violence has focused extensively on identifying individual-level risk factors to inform its prevention and intervention. A significant line of research has centered on the application of psychological theories to elucidate cognitive, behavioral, and situational factors that elevate sexual violence perpetration risk. These empirical efforts have culminated in the development of comprehensive theoretical frameworks, most notably the Confluence Model (Malamuth et al., 1991, 1995). As one of the most extensively validated and influential theoretical paradigms in the field, the Confluence Model explicates the complex interaction of multiple risk factors in male-perpetrated sexual aggression against women. It posits that sexual aggression emerges from the convergence of two primary pathways, Hostile Masculinity (HM) and Impersonal Sex (IS), along with a series of distal and contextual/situational risk factors, such as delinquency, psychopathy, lack of empathy and alcohol consumption. The two primary risk factors, HM, characterized by gender-based violent and distrustful cognition toward women, and IS, reflecting a casual, non-committal approach to sexual relationships, not only interact individually, but synergistically, to increase risk of sexual aggression (Huntington et al., 2022; Saqib & Davidson, 2023). However, despite three decades of rigorous investigation, certain structural limitations of the Confluence Model warrant attention. In particular, the operationalization and measurement of HM lack standardization, and its dimensional structure has not been comprehensively evaluated. Accordingly, the present study proposes a comprehensive psychometric analysis of HM within the Confluence Model framework to verify its dimensionality and establish standardized measurement protocols for future research.

Manuscript Organization

The manuscript is structured around a three-study design aimed at refining the conceptualization and measurement of HM within the Confluence Model of sexual aggression. It begins with a scoping review of the Confluence Model, emphasizing the theoretical and empirical development of HM and identifying existing gaps in measurement consistency and dimensionality. This is followed by a summary of the project's theoretical foundation and rationale, which leads into the identification of four core research objectives that form the basis for the three-study design. The subsequent section presents the three studies in sequence. Each study begins with a description of the statistical analysis plan, including justifications for selected data analysis tools, software, and model fit indices. This is followed by the methodological section, which provides comprehensive protocols describing sampling strategies, study procedures, and measures applied. The results and discussion for each study are then presented. The manuscript concludes with a general discussion that integrates the findings across all three studies and highlights their broader theoretical and methodological implications, as well as limitations and directions for future research.

Literature Review

The Confluence Model of Sexual Aggression

The Confluence Model is rooted in the work of Malamuth (1986), which sought to integrate previously isolated risk factors into a more coherent theoretical and empirical framework for understanding and predicting male-perpetrated sexual aggression against women. The model's initial formulation synthesized and evolved insights from several influential theories, including feminist theory (Brownmiller, 1993), evolutionary theory (Buss & Schmitt, 1993), and social learning theory of aggression (Bandura, 1978). For example, the feminist perspectives framed sexual aggression as an act of power, control, and dominance embedded in patriarchal social structures, which serves to maintain a climate of fear and to reinforce gender inequalities. These perspectives foregrounded the initial conceptualization of HM, which reflects the socialization of men into attitudes supporting violence against women. Evolutionary theory explained sexual aggression as a byproduct of evolved mating strategies that arise from conflicts between male and female reproductive interests. Because male reproductive interests are maximized by seeking multiple partners while female interests are maximized by selectivity and resource investment, males tend to be more oriented toward engaging in casual, uncommitted sexual relationships (i.e., sociosexuality; Penke & Asendorpf, 2008). This concept was later incorporated into the Confluence Model as IS (Malamuth et al., 1995). Finally, insights from social learning theory of aggression (Bandura, 1978), particularly its extended application to childhood sexual abuse (Finkelhor & Araji, 1986) and sexual aggression (Russell, 1984), also informed the early development of the Confluence model, underscoring that sexual aggression arises from multiple, interacting factors rather than from any single cause. For example, Finkelhor and Araji's (1986) four-factor model proposed that child sexual abuse arises from four interrelated processes including emotional congruence (viewing children as emotionally compatible partners), sexual arousal to children, blockage of adult peer sexual outlets, and disinhibition from conventional restrictions. Each of these factors has both individual and social/cultural levels of explanation. Several elements of these explanations, such as childhood victimization, pornography exposure, alcohol consumption, psychosis, and lack of empathy, are later identified as secondary risk factors in subsequent refinements of the Confluence Model (Malamuth et al., 2012, 2021; Parkhill & Abbey, 2008; Wheeler et al., 2002). Similarly, Russell (1984) underscored that sexual aggression is socially learned, through not only interpersonal modeling but also systemic cultural reinforcement of male power and rape-tolerant norms. It is deeply embedded within a patriarchal cultural framework, which teaches, tolerates, and reinforces male dominance and aggression toward women and children. These norms legitimize aggression in sexual relationships and are internalized by young men, particularly when reinforced by peers, pornography, and societal silence around abuse.

Building on these insights, Malamuth developed the Confluence Model, aiming to identify characteristics of men at heightened risk of perpetrating sexual aggression, including acts that meet legal definitions of rape and are seldom prosecuted by the justice system (Malamuth & Hald, 2016). The model emphasizes a hierarchical-mediational structure involving two central, interactive constellations or aggregative characteristics:

HM and IS. These two proximal risk factors mediate a series of secondary influences, including both general antisocial characteristics and risk factors specific to sexual aggression, in predicting sexual aggression (Malamuth et al., 2021).

HM refers to “a set of traits that are associated with an insecure, defensive, distrustful, hostile and domineering orientation towards women (Ray & Parkhill, 2021, p.356).” The construct was first conceptualized to represent two of the most primary drives of sexual aggression beyond sexual gratification: rape for power and rape for anger (Malamuth, 1986). Research has indicated that these two drives are largely affected by the cultural notion of masculinity, which emphasizes power, dominance and toughness (Connell & Messerschmidt, 2005; Levant et al., 2010). Men who strongly adopt these masculine cultural norms may experience anxiety and insecurity about adequately fulfilling such expectations, potentially leading to compensatory behaviors that overtly display masculine characteristics, especially when their masculine identity feels uncertain or challenged (Vandello & Bosson, 2013). In sexual relationships or intimate contexts, sexual aggression may serve as a mechanism through which men establish dominance, assert control, and consolidate power (Groth & Birnbaum, 2013; Zurbriggen, 2000). As such, the HM Path to sexual aggression refers to “a personality profile combining two interrelated components: (a) a narcissistic, insecure, defensive, hypersensitive, and hostile-distrustful orientation, particularly towards women, and (b) sexual gratification from controlling or dominating women (Malamuth & Hald, 2016, P.54).” This path leading the HM constellation also contains a series of antecedents of HM, such as supportive attitudes toward interpersonal violence, especially violence against women, and personality traits associated with general antisocial behaviors (e.g., psychopathy).

IS, as the other central risk factor of the Confluence Model, was first introduced by Malamuth et al. (1995) as a construct closely related to sociosexuality, an evolutionary concept broadly defined as an individual's willingness to engage in uncommitted sexual relations (Simpson & Gangestad, 1991). It reflects one's emotional detachment from sexual relationships and a preference for short-term and casual sex over building intimate relationships and long-term commitments with sexual partners. This impersonal orientation to sex was theorized to develop in the context of growing up in an abusive, violent or neglectful home environment, which impairs the ability to establish healthy attachment, emotional regulation and interpersonal skills that are typically socialized through early parent-child interactions. These early adverse childhood experiences, in turn, increase the vulnerability to antisocial or delinquent behaviors during adolescence and to the adoption of delinquency-supportive norms from peers. Individuals exhibiting high IS often view sexuality as a game to be won and a source of peer status and self-esteem, which in turn encourages their engagement in promiscuous sex and the use of various coercive tactics to obtain sex at any cost. This IS path was originally labeled as the Sexual Promiscuity/Impersonal Sex path, indicating both a history of casual sexual experiences and ongoing desires and fantasies about uncommitted sex partners, such as strangers, prostitutes, or extra-relationship partners (Malamuth et al., 1995; Malamuth et al., 1996; Vega & Malamuth, 2007). The term sexual promiscuity was phased out in recent research in favor of more neutral "IS" to avoid potential moral judgment

(Malamuth et al., 1996) as compared to the HM constellation, which primarily reflects personality traits; the IS constellation consists more of experiential/behavioral ones, including childhood victimization and adolescent delinquency (Malamuth & Hald, 2016).

Beyond HM and IS, the Confluence Model also posits a broader set of distal, general antisocial factors that shape individuals' developmental trajectories and feed into these two proximal pathways, which in turn predict sexual aggression. From its earliest developments, the model has emphasized the role of adverse childhood and adolescent experiences, such as growing up in abusive, neglectful, or hostile family environments, witnessing interparental violence, experiencing childhood physical or sexual abuse, and engaging in adolescent delinquency, in disrupting the development of secure attachment, healthy relationship functioning, empathy and emotion regulation (Malamuth et al., 1991, 1995). These developmental disruptions increase the likelihood of affiliating with delinquent peers in order to meet unmet needs for belonging, validation, and social support that are lacking in early caregiving environments, which promotes the adoption of pro-delinquent peer norms and aggressive interpersonal attitudes that foster a broader pattern of antisocial behavior that sets the stage for later sexual aggression. Within the Confluence Model, these distal experiences are theorized to influence sexual aggression primarily through mediation by the "specialized" proximal risk factors of HM and IS. When a child grows up in an abusive, violent, and neglectful home environment, chronic exposure to hostility and instability can impair the development of secure attachment, emotion regulation, and trust in intimate relationships. As a result, the child may learn to view interpersonal interactions, particularly intimate relationships, through a lens of dominance, hostility, and mistrust. Over time, these violence-oriented gendered cognitions increase difficulties in forming emotionally close and reciprocal relationships while simultaneously encouraging an impersonal orientation toward sexual relationships characterized by casual sex, emotional detachment, and sexual opportunism. Empirical tests of the model have supported these pathways, demonstrating that delinquency is associated with sexual aggression both directly and indirectly via HM and IS, whereas childhood victimization more consistently operates through HM (Abbey et al., 2011; Malamuth et al., 1991, 1995). Importantly, the strength and structure of these pathways appear to depend on how adverse childhood experiences are operationalized. When measured broadly to include emotional, physical, and sexual abuse, childhood adversity is conceptualized as contributing to generalized antisocial tendencies, with sexual aggression emerging as one manifestation among a broader constellation of aggressive behaviors. In contrast, when childhood sexual abuse is assessed specifically, it often shows more direct associations with later sexual aggression, bypassing general delinquency pathways and reflecting social learning processes through which patterns of sexual coercion are internalized and later reenacted (Bandura, 1978; Loh & Gidycz, 2006; Milaniak & Widom, 2015; Nguyen & Parkhill, 2014).

In addition to the distal developmental antecedents, the Confluence Model was later expanded to incorporate a cluster of secondary or situational/contextual factors, including pornography use (particularly violent or degrading porns; Malamuth et al., 2000), deficits in emotional regulation, especially impulse-control difficulties (Kirwan et al., 2019), low empathy (Wheeler et al., 2002), peer support for aggression (Malamuth et al., 2021), and alcohol consumption (Abbey et al., 2006; Parkhill & Abbey, 2008; Rabaud et al., 2025; L. Smith et al., 2024). Specifically, alcohol consumption here has been conceptualized both as a general pattern of frequent consumption and as situational drinking in high-risk contexts, such as drinking parties, dating encounters, or sexual situations (See Rabaud et al., 2025 for review). These secondary risk factors function primarily as moderators that do not independently cause sexual aggression, but interact with the proximal risk factors to elevate risk. For example, extreme pornography use has been described as "fuel in the fire" (Malamuth et al., 2012, p.429), in that it heightens the likelihood of sexual aggression mainly among men already high in levels of HM and IS, rather than initiating aggression among low-risk individuals. Similarly, low empathy and poor impulse control weaken internal inhibitions against harming others, while peer norms supportive of aggression and alcohol-related disinhibition further increase the probability of at-the-moment sexual aggressive impulse to be enacted.

Together, the Confluence Model is best characterized as a mediated-moderation framework in which distal antisocial experiences shape proximal risk factors and secondary contextual or situational influences determine whether and when those risks translate into sexually aggressive behaviors. The perpetration of sexual aggression is therefore most likely to occur when multiple risk factors converge, particularly among individuals high in HM and IS who frequently participate in environments that facilitate disinhibition, objectification, and social reinforcement of aggression.

The Outcome Measure of Sexual Aggression

The model operationalizes its outcome, sexual aggression, using the Sexual Experiences Survey-Perpetration (SES-P) or its modified versions (Abbey et al., 2005, 2021; Abbey & Jacques-Tiura, 2011; Koss et al., 2007; Koss & Gidycz, 1985; Lisak & Roth, 1988) as the gold standard outcome measure. First introduced by Koss and her colleagues (1982), the SES was designed to assess both victimization and perpetration through self-report, focusing on behaviorally specific acts rather than legal definitions of sexual assault or rape. Given that a large number of sexual assaults go unreported, the SES conceptualized sexual aggression along a continuum encompassing behaviors from psychological, verbal coercion to attempted or completed rape, thereby sensitively capturing sexual aggressive behaviors that often remain undetected by the justice system (Abbey et al., 1998; Koss, 1996). Many early Confluence Model studies (Anderson & Anderson, 2008; Calhoun et al., 1997; Lalumière et al., 1996; Malamuth et al., 1995) adopted Koss et al.'s (1987) version of SES, which assessed ten items, each asking about a sexual act (e.g., oral-genital contact, attempt or completed genital/anal intercourse) paired with the tactic used to obtain it. A sample item states: "Have you had sexual intercourse (i.e., act) when the woman didn't want to because you threatened or used some degree of physical force (i.e., tactic) to make her?"

Over time, the measure has undergone several notable revisions, which were incorporated into the subsequent testing of the Confluence Model. For example, some studies employed a modified 16-item scale (Abbey et al., 2006, 2011) that added six items to the original 10-item version (Koss et al., 1987), providing more specific examples of incidents such as verbally coerced oral and anal sex induced through swearing, anger, or threats, as well as taking advantage of a woman who had passed out or was too intoxicated. Other studies employed Abbey et al.'s (2005) 35-item version which assessed five types of forced sex (i.e., "sexual touching, attempted intercourse, oral sex, completed sexual intercourse, and anal sex/insertion of objects") and seven tactics (i.e., "using arguments and pressure, telling lies and making promises, using guilt and anger, giving alcohol, giving drugs, taking advantage of a woman who is too intoxicated to consent, and using physical force", p. 5). Compared to the original SES (Koss et al., 1987), this modified version expanded the scope of the scale and adjusted the question phrasing by presenting the tactics used as an initial frame of reference, followed by the specific acts associated with those tactics. This decision was guided by the findings that a tactic-first question frame yielded a significantly higher endorsement rate than the version that presented acts first. Koss and her colleagues (2007) also released a major revision of the short form of the SES (SES-SFP). In this version, several changes were made to improve the behavioral specificity, gender neutrality, and item clarity. For example, the act of sexual touching was further specified as acts against the private areas of the body, including lips, breast/chest, crotch or butt. The SES-SFP, like Abbey et al.'s (2005) modified version, included 35 items covering seven types of forced acts and five tactics, but retained the acts-first question frame, emphasizing more detailed and behaviorally explicit phrasing.

Studies applying different versions of the SES adopted diverse scaling and scoring methods. The scaling approaches ranged from the original dichotomous Yes/No format to 4- to 7-point Likert scales, with some studies employing relatively open-ended numerical ranges (e.g., 0 to 20+). Concerning scoring methods, some studies directly summed these responses to create a total score, while others employed weighted scoring methods to capture variability in the severity and types of aggression perpetrated (Abbey et al., 2011; Casey et al., 2017). Davis et al. (2014) compared different scoring methods for the SES and found that each method was viable, yielded comparable results, and had a strong association with convergent measures. Yet, it was also noted that certain methods (i.e., accounting for multiple types and frequencies of assault) offer distinct advantages in samples with greater perpetration rates.

The Conceptualization and Measurement of Hostile Masculinity (HM)

Despite around three decades of empirical testing of the connection between HM and male-perpetrated sexual aggression, substantial conceptual and measurement inconsistencies remain in the literature. These inconsistencies originate in the construct's historical development and have persisted as the Confluence Model has been extended and operationalized across studies. From its early formulation, the construct of HM did not emerge as a unidimensional latent variable. Rather, it evolved from Malamuth's (1986) interactive model of sexual aggression, which identified multiple motivational and disinhibitory predictors that operated synergistically to elevate risk of perpetration. In this initial model, sexual aggression was theorized to arise from the convergence of distinct motivational drives, including rape for power, rape for anger, and rape for satisfying sexual desires and gratification, alongside factors that lower internal and external inhibitions. These motivations were operationalized using conceptually distinct measures, including sexual dominance (Nelson, 1979), hostility towards women (Check & Malamuth, 1983) and sexual responsiveness to rape. Additional predictors, such as attitudes facilitating aggression (Hurt, 1980) and antisocial personality traits (Eysenck & Eysenck, 1978), were incorporated to capture broader dispositional and attitudinal disinhibitors that reduce deterrence and moral restraint.

Some of these components, originally modeled as separate predictors, including sexual dominance, hostility toward women, and attitudes supporting violence, were later integrated to define HM. Specifically, sexual dominance reflected a desire to conquer and exert power over women; hostility toward women captured persistent distrust and antagonism toward women; and acceptance of interpersonal violence addressed beliefs that the use of force was an acceptable strategy to establish compliance, particularly in intimate contexts. Although the initial work (Malamuth, 1986) also considered rape myth acceptance (i.e., false and stereotypical beliefs about rape; Burt, 1980; Lonsway & Fitzgerald, 1995) and the adversarial sexual beliefs (i.e., notions that male-female relationships are inherently exploitative and manipulative; Burt, 1980), constructs that were commonly incorporated in later studies, they were found to be less central to explaining sexual aggression than acceptance of interpersonal violence, and therefore were not emphasized in the early model. Collectively, these components laid the conceptual groundwork for HM as a dispositional orientation characterized by an insecure, defensive, distrustful, hostile and domineering cognition toward women.

As the Confluence Model matured, HM emerged as one of the two central, proximal pathways to sexual aggression, often operationalized as a multidimensional, composite latent construct. However, the specific components used to define HM have varied considerably across studies. As documented in the systematic review by Ray and Parkhill (2023), researchers have inconsistently combined hostility toward women, adversarial sexual beliefs, sexual dominance, rape myth acceptance, and acceptance of interpersonal violence when operationalizing HM, with no clear consensus regarding which components should be considered core versus secondary. The core components most commonly used to define HM include hostility toward women, adversarial sexual beliefs, and sexual dominance. In contrast, rape myth acceptance and acceptance of interpersonal violence have been incorporated less consistently. In some studies, these latter constructs were included within HM composites.

In contrast, in others, including the original work of Malamuth et al. (1991, 1995), they were treated as conceptually related but distinct attitudinal variables, as an early etiological “attitudes supporting violence” construct along the HM pathway. The review emphasized that variability in the components' inclusion has generated ambiguity in construct definition and interpretation. Although HM composites, regardless of their specific composition, have shown robust associations with sexual aggression, inconsistent operationalization complicates interpretation and limits comparison across studies. Such variability may obscure whether observed associations reflect a coherent underlying construct or a convergence of overlapping but distinct attitudinal and motivational processes. To address these concerns, Ray and Parkhill (2023) recommended standardizing the conceptualization of HM by defining it through hostility toward women, adversarial sexual beliefs and sexual dominance, while rape myth acceptance and acceptance of interpersonal violence as a conceptually related but separate attitudes supporting violence (ASV) construct.

Even though the recommendation offers a theoretically informed framework for operationalizing HM, rigorous evaluation of the construct's measurement properties and dimensionality remains lacking in the literature. Past empirical research has most often operationalized HM either as a composite score derived from multiple subscales or as a single latent factor within applications of the Confluence Model, without explicitly evaluating whether the assumption of unidimensionality is empirically supported by the item-level measurement structure. In some cases, item parceling has been employed to simplify model estimation, often in response to limitations in available statistical software or computational capacity at the time, which also obscured potential multidimensionality by aggregating items across theoretically distinct components (Dean & Malamuth, 1997; Malamuth et al., 1995, 2021; Parkhill & Abbey, 2008; Saqib & Davidson, 2023). These practices implicitly assume unidimensionality and equivalence among components, despite conceptual distinctions among hostility among women, adversarial sexual beliefs and sexual dominance. These limitations are particularly consequential given that the Confluence model is developed using a structural equation modeling framework (Malamuth et al., 1991), in which valid structural inferences depend critically on the adequacy of the underlying measurement models.

Structural Equation Modeling and Item Parceling

Structural Equation Modeling (SEM) is a well-established statistical modeling framework that has been widely applied to construct and test psychological models

(MacCallum & Austin, 2000; J. Ullman, 2006). SEM was first applied by Malamuth et al. (1991) for developing and testing the Confluence model and has been continuously used in subsequent model examinations (Brazil et al., 2023; Davis & Logan-Greene, 2012; Dean & Malamuth, 1997; Kirwan et al., 2025; Malamuth et al., 2021; Yucel et al., 2023). SEM is often used to estimate latent constructs using multiple observed indicators and allows simultaneous examination of direct and indirect pathways between developmental antecedents and sexually aggressive behavior. SEM with latent factors (hereafter referred to simply as SEM) benefits the Confluence Model construction in several ways (Malamuth & Hald, 2016). First, SEM allows the modeling of measurement error, which improves the precision of model estimates by correcting for variability and bias in latent factor estimates. Second, SEM enables the statistical testing of the risk factor grouping/organization of the Confluence Model according to their shared variance. For example, the personality and cognitive traits related to psychopathy were modeled as one pillar (i.e., broader construct) that predicts general antisocial behaviors. Third, SEM's path analytic testing also enables the modeling of direct relationships between specific factors (such as childhood victimization) and sexual aggression, as well as their indirect effects through mediation and/or moderation with more proximal risk factors (e.g., IS).

To achieve this, a two-step approach is generally recommended for the testing of SEM (Anderson & Gerbing, 1988), which involves: (1) estimating and validating the measurement model and (2) estimating the structural model, which tests the hypothesized theoretical paths. The examination of the measurement model is typically conducted using Confirmatory Factor Analysis (CFA; Brown, 2015) to examine the dimensionality and factor structure by separating how observed variables (indicators) relate to their underlying latent variables (Anderson & Gerbing, 1988; MacCallum & Austin, 2000). In other words, CFA evaluates whether the observed items load appropriately onto the hypothesized factors. If CFA supports the hypothesized structure, additional scale refinement may involve evaluating item reliability and discriminant validity. If not, the scale may require revision: this could include omitting items with high unique variance or complex loadings, or developing items to better represent the intended constructs. This reassessment of even previously validated measures is needed prior to the simultaneous estimation of the measurement and structural models, also because it allows for a confirmatory assessment of discriminant validity across all latent constructs that are intended to be part of the same structural model. When two constructs are too highly correlated, they might not be sufficiently separable. For example, if the HM and attitudes supporting violence constructs are not different enough, it may raise concerns that one of these constructs is redundant, misspecified, or poorly operationalized.

Although SEM offers the advantage of modeling latent variables while accounting for measurement error, many applied studies do not fully utilize this strength. Instead, they rely on path analysis with observed variables, which can lead to biased parameter estimates. As noted by MacCallum and Austin (2000), more than half of the early articles published reported path analyses involving only observed variables. Ignoring or underreporting of the measurement model persists in more recent studies, particularly through the improper use of item parceling. Item parceling is commonly employed to reduce model complexity in SEM, which involves aggregating subsets of indicators (e.g., summing or averaging 3 to 4 items) to create multiple parcels. These parcels are then used as observed indicators of a latent variable in structural model testing (Bandalos, 2002; Hall et al., 1999; Matsunaga, 2008). There are two common strategies for parceling, including distributed parceling and isolated parceling (Bandalos, 2008; Hall et al., 1999; Rhemtulla, 2016). In distributed parceling, items from different subdimensions are evenly mixed across parcels to emphasize the general factor and improve model fit. Isolated parceling, by contrast, groups items loading on the same factor into separate parcels, allowing for a clear representation of theoretically or empirically supported subdimensions of a construct. Either strategy helps to reduce the number of estimated parameters, which in turn increases degrees of freedom, stabilizes parameter estimates, and often improves overall model fit (Little et al., 2002).

Item parceling is considered beneficial, particularly when dealing with non-normal data, small sample size, or less precise scales with few response options (Bandalos, 2008; Little et al., 2002). However, a critical assumption of item parceling is that the items being grouped to form parcels collectively reflect a single, unidimensional construct (Bandalos, 2002; Sass & Smith, 2006). When this assumption of unidimensionality is violated, parceling can obscure the true multidimensional structure of the data, failing to include secondary factors and potentially leading to biased estimates of structural paths (Hall et al., 1999). In the context of HM, for instance, items may reflect distinct subdimensions/scales, such as hostility toward women (HTW), adversarial sexual beliefs (ASB), and Sexual Dominance (SD). If a distributed parceling strategy is applied, items from different subdimensions will be mixed within each parcel (e.g., Parcel A: 4 HTM, 4 ASB; Parcel B: 4 ASB, 4 SD; Parcel C: 4 HTM, 4 SD). The variance attributable to potential secondary dimensions might be inappropriately absorbed into the general HM factor, thereby biasing model fit statistics.

Additionally, if a secondary dimension like ASB shares unique variance with another variable in the model (e.g., IS, sexual aggression), the structural path from HM to that outcome might be biased, thereby distorting the estimated relationships among latent variables. Alternatively, if an isolated parceling strategy is used, which maintains parcels by subdimensions (e.g., Parcel A: ASB, Parcel B: HTW, Parcel C: SD), then any unique variance associated with each subdimension will not be modeled directly. As a result, any differential variance explained by ASB, HTW, or SOD to outcomes will be treated as measurement error in the model, potentially attenuating the estimated effects of HM and biasing the overall model fit.

Given the importance of evaluating the unidimensionality assumption, factor analysis is generally recommended before item parceling (Bandalos, 2002; Kishton & Widaman, 1994, 1994; Sass & Smith, 2006). However, many studies continue to bypass this essential step in practice (Coffman & MacCallum, 2005; Hall et al., 1999; Little et al., 2002; Marsh et al., 2013). Bandalos and Finney (2001) reported that only 32.3% of studies using item parceling explicitly tested the unidimensionality of the items being parceled, which raised concerns about the validity of their structural conclusions. Sterba (2019) also acknowledged that item parceling has frequently been applied to constructs that are theoretically multidimensional, raising concerns of parcel allocation variability, which can result in unstable and inconsistent parameter estimates across different parcel allocations.

Although the Confluence Model was initially developed using the SEM framework due to its strengths in modeling complex interrelationships among latent constructs (Malamuth et al., 1991; Malamuth et al., 1995), few subsequent studies have conducted a thorough SEM examination. Instead, many studies chose path analysis or decomposed paths into multiple regression models. For example, Malamuth et al. (2000) explained their use of path analysis rather than full SEM due to limited sample size constraints. Parkhill and Abbey (2008) opted for path analysis because most of the factors examined contained only one or two indicators, making a full SEM with the measurement component inappropriate. Wheeler et al. (2002) and Wright et al. (2021) conducted a series of hierarchical regressions because they focused on a single secondary risk factor (i.e., empathy and pornography exposure, respectively) and examined how they interact with the primary risk factors (HM and/or IS) to predict a higher level of sexual aggression. Other researchers have claimed to apply SEM but failed to report detailed measurement model analyses (Brazil et al., 2023; Davis & Logan-Greene, 2012; Dean & Malamuth, 1997; Malamuth et al., 2021). It is possible that some studies conducted factor analyses but did not report them because of model complexity or manuscript space limitations, whereas others may have relied on prior scale validation without conducting additional factor analyses. Studies also do not always specify whether factors were represented using composite scores or item parcels. When parceling was used, relatively few provide details about parcel allocation procedures or report dimensionality checks used to justify those decisions (Yucel et al., 2023). This inconsistency and frequent lack of clarity in measurement validation reporting is particularly problematic for theoretically multidimensional constructs such as HM. In fact, few published Confluence Model studies to date appear to have conducted a comprehensive factor analysis or dimensionality assessment of HM. Careful and explicit measurement model testing and validation remain essential prerequisites for understanding the internal structure and psychometric properties of HM when considering item parceling or latent construct modeling approaches for the Confluence Model.

Proposed Research

Overview of Objectives and Data Analysis Plan

The theoretical foundation of this study rests on the understanding that HM is an inherently multidimensional construct. Previous research has often used item parceling to model HM as a unidimensional construct, which may obscure important nuances among its underlying dimensions. Although item parceling is a defensible strategy for simplifying the structural model when a construct is truly unidimensional, a thorough psychometric evaluation is needed to empirically determine the dimensional structure that best represents HM. To address this gap, the current project combined existing datasets using an Integrative Data Analysis (IDA) framework. At the item level, graded response item response theory models (GRM-IRT) were used to evaluate the functioning of individual HM subscale items and to examine differential item functioning (DIF) across studies, assessing between-study heterogeneity and potential sources of measurement bias. At the construct level, bifactor S-1 ESEM were used to test the multidimensional structure of HM and evaluate whether a general HM factor, along with more specific subfactors, provides the best representation of the data. Finally, using newly collected data from CloudResearch, the revised HM scale was validated and situated within a broader nomological network of the Confluence Model risk factors. With this in mind, the following objectives were specified to guide this project:

(1) To conduct item-level psychometric evaluation of HM subscales using IRT across multiple existing datasets (Study 1; Dataset 1: N=229 & Dataset 2: N=491)

(2) To examine DIF to assess between-study heterogeneity and measurement invariance before integrating datasets (Study 1; Dataset 1: N=229 & Dataset 2: N=491)

(3) To evaluate the dimensionality of the HM construct using bifactor S-1 ESEM and competing measurement models with the integrated dataset (Study 2; IDA: N=720)

(4) To explore and establish evidence for a nomological network of HM scale scores, utilizing empirically validated measures of other Confluence Model risk factors for sexual aggression (Study 3; CloudResearch: N=329)

Studies 1 and 2 involved secondary analyses of archived data from previously conducted studies. Study 3 procedures received ethical approval from the Oakland University Institutional Review Board before their implementation (see Appendix A).

CHAPTER TWO

Study 1 (S1)

To increase sample size and evaluate item-level measurement structure and between-study heterogeneity in HM, the current study employed an IDA framework to combine data from two independent studies. IDA is a methodological framework that enables the simultaneous analysis of raw data pooled from multiple independent studies (Curran & Hussong, 2009; Hussong et al., 2013). It is particularly advantageous for research programs as it allows investigators to increase statistical power, examine cross-study heterogeneity, and evaluate measurement structure in ways that are not possible within any single dataset. Unlike meta-analysis, which synthesizes summary statistics from completed studies, IDA operates at the level of raw participant data. This allows for more precise and flexible modeling, greater power, and improved model stability when estimating complicated models (Curran et al., 2008). Moreover, item-level pooling also enables the direct examination of sample heterogeneity to carefully select the contributing studies with minimum difference (Hussong et al., 2013). If any between-study heterogeneity is detected, IDA also holds the potential to distinguish between within- and between-study sources of variation and determine whether theoretically meaningful effects replicate across studies.

It should be noted that IDA is not a specific technique or analysis, but rather a flexible framework that guides the design, measurement, and modeling decisions involved in pooled data analysis (Hussong et al., 2013). A central component of IDA is the construction of harmonized items and commensurate measures. Harmonized items refer to items that have been transformed or recorded so that they are directly comparable across studies, while commensurate measures involve the use of psychometric models to ensure the resulting scales have the same meaning and metric across studies. This process acknowledges that the identical items may function differently across populations or study contexts, and thus requires explicit evaluation of measurement equivalence (Bauer & Hussong, 2009). Following Hussong and colleagues (2013), the IDA typically proceeds through four iterative steps, including preliminary feasibility analysis, item selection, measurement model development, and scoring for final model evaluation. The first step of preliminary feasibility analysis involves determining whether pooling is justified by examining the comparability of samples and assessing whether studies include overlapping constructs or common items that can link measurement across datasets. After establishing feasibility, the step of item selection involves identifying item sets based on theoretical relevance and available measurement content. Descriptive analyses and dimensionality assessments are also conducted in this step to evaluate item quality, detect potential multidimensionality, and identify problematic items. The third step pertains to measurement model development, estimating a psychometric model, often an IRT or similar latent variable model, to examine item functioning and test for Differential Item Functioning (DIF) across studies. When necessary, test score linking procedures are applied to place all participants on a common latent metric. Finally, scoring and model evaluation involve generating latent scores from the calibrated model while incorporating detected DIF or impact parameters to correct for potential sources of bias. This process acknowledges that even identically worded items may function differently across samples or study contexts and therefore necessitates explicit evaluation of measurement equivalence before pooled inference.

Following the established IDA guidelines (Hussong et al., 2013), the current study applied this framework to merge, analyze, and compare information from two contributing datasets. The codebooks of each dataset were already reviewed to ensure they include the HM subscales (i.e., SOD, HTW, and ASB). IRT and DIF analyses were used for IDA to prepare the pooled item-level data for subsequent latent construct modeling by evaluating item functioning, detecting potential between-study heterogeneity, and generating commensurate HM subscale scores.

Method

Participants

The current study adopted data drawn from two existing studies for IDA:

Project Become 1. Project Become 1 was a cross-sectional study that examined dating and sexual experiences conducted in 2017. Samples for this dataset were collected at a large suburban Midwestern university through the online subject pool of the psychology department, administered via Qualtrics using SONA. Eligibility criteria required participants to be at least 18 years old and enrolled in one of the lower-division psychology courses. The final sample consisted of 229 male undergraduate students ranging in age from 18 to 40 years old (M_{age} = 21.45, *SD* = 4.67). Participants primarily identified as White/Caucasian (70.7%), with additional representation from African American (7.9%), Middle Eastern (5.7%), Asian American (5.7%), Native American (0.4%), and Pacific Islander (0.4%) backgrounds; 2.2% identified as hispanic/latino(a), 2.2% as “other,” and an additional 2.6% declined to report their race. Regarding sexual experiences, the majority of participants reported exclusively heterosexual experiences (91.7%), with smaller proportions reporting mostly heterosexual experiences (2.3%), bisexual experiences (1.5%), or exclusively homosexual experiences (1.5%); 3.0% declined to respond. See Table 1 for participant demographic characteristics of Project Become 1.

Table 1
Demographics for Project BECOME 1 Participants (N = 229)

Demographic Variables		n	%
Age	18-24	198	86.50 %
	25-29	20	8.70%
	30-39	10	4.40%
	40 or older	1	0.40%
Race/Ethnicity	European American/White/Caucasian	162	70.70 %
	African American/Black	18	7.90%
	Asian American	13	5.70%
	Middle Eastern/Arabic	13	5.70%
	Hispanic/Latino(a)	5	2.20%
	Multiracial	5	2.20%
	American Indian/Alaska Native	1	0.40%
	Native Hawaiian/other Pacific Islander	1	0.40%
	Histpanic/Latino(a)	5	2.20%
	Other	5	2.20%
	Prefer not to answer	6	2.60%
Class Standing	Freshman	54	23.60 %
	Sophomore	70	30.60 %
	Junior	58	25.30 %
	Senior	38	16.60 %
	Graduate student	1	0.40%
	Other	6	2.60%
	Prefer not to answer	2	0.90%
Sexual Experiences	Exclusively heterosexual	214	93.40 %
	Mostly heterosexual with some homosexual experience	5	2.20%
	Bisexual	2	0.90%
	Mostly homosexual with some heterosexual experience	2	0.90%
	Exclusively homosexual	6	2.60%

Note. N = 229. Percentages are based on the full Project BECOME 1 sample. Age ranged from 18 to 45 years (M = 21.21, SD = 4.02). Open-ended "other" responses are collapsed into their respective "Other" categories.

Project Become 2. Project Become 2 is a cross-sectional online study conducted in 2019. Samples for this dataset were collected via Amazon Mechanical Turk (MTurk) between January and July 2021 as part of an online study examining men's dating and sexual experiences. Participants completed a 45-minute self-report questionnaire administered through Qualtrics. Eligibility criteria required participants to reside in the United States, be at least 18 years old, identify as male, and have a minimum 95% prior task approval rate, consistent with best practices for MTurk research. CAPTCHA verification and attention checks were embedded to ensure data quality. Participants received $3.00 in compensation upon completion. The final sample consisted of 491 adult men (M_{age} = 30.52, *SD* = 8.85). Participants primarily identified as White/European American (63.5%), with additional representation from Black/African American (11.0%), Hispanic/Latino (11.0%), and Asian American or Pacific Islander (8.8%) individuals. Most participants reported at least some postsecondary education, with approximately half holding a bachelor's degree or higher (49.7%). The majority reported an annual household income of $50,000 or above (67.8%). Regarding sexual orientation, most participants identified as heterosexual (86.8%), followed by bisexual (10.2%), with a small proportion identifying as another sexual identity (3.1%). See Table 2 for participant demographic characteristics of Project Become 2.

Table 2
Demographics for Project BECOME 2 Participants (N = 491)

Demographic Variables		n	%
Age	18-24	139	28.30%
	25-29	116	23.60%
	30-39	155	31.60%
	40-49	61	12.40%
	50 or older	19	3.90%
	Missing/uncodable	1	0.20%
Race/Ethnicity	European American/White	312	63.50%
	African American/Black	54	11.00%
	Hispanic/Latino(a)	54	11.00%
	Asian American	41	8.40%
	Multiracial	13	2.60%
	American Indian/Alaska Native	7	1.40%
	Middle Eastern/Arabic	6	1.20%
	Native Hawaiian/other Pacific Islander	2	0.40%
	Not listed	1	0.20%
	Missing	1	0.20%
Sexual Orientation	Heterosexual/straight	426	86.80%
	Bisexual	50	10.20%
	Gay	8	1.60%
	Other identity	7	1.40%
Sexual Experiences	Exclusively straight	391	79.60%
	Mostly straight with some gay experience	48	9.80%
	Bisexual	43	8.80%
	Mostly gay with some straight experience	9	1.80%
Education	Less than high school diploma	4	0.80%
	High school diploma	58	11.80%
	GED or ABE certificate	8	1.60%
	Some college	126	25.70%
	Associate's/technical degree	51	10.40%
	Bachelor's degree	176	35.80%
	Graduate degree	68	13.80%
Family Yearly Income	Under $9,999	9	1.80%
	$10,000-$19,999	35	7.10%
	$20,000-$29,999	34	6.90%
	$30,000-$39,999	39	7.90%
	$40,000-$49,999	56	11.40%

	$50,000-$59,999	62	12.60%
	$60,000-$69,999	34	6.90%
	$70,000-$79,999	65	13.20%
	$80,000 or more	157	32.00%
U.S. Region	West	101	20.60%
	Southwest	51	10.40%
	Midwest	102	20.80%
	Southeast	121	24.60%
	Northeast	116	23.60%

Notes. Percentages are based on the full Project BECOME 2 sample. One participant entered “20’s” for age and was treated as missing/uncodable for the grouped age categories. Open-ended “other” responses were collapsed into their respective “Other” categories.

Final Pooled Sample. The final pooled sample consisted of a total of 720 participants, 229 of which were drawn from Project Become 1 (age ranging from 18 to 40 years) and 491 from Project Become 2 (age ranging from 18 to 66 years). All participants self-identified as men, and most identified as heterosexual or mostly heterosexual. This sample size exceeds commonly recommended thresholds for estimation of GRM with Likert-scale data. Simulation and empirical studies indicated that stable item parameter recovery in GRM is generally achieved with samples of at least 300 respondents, particularly when scales include five or more items (Kieftenbeld & Natesan, 2012; Reise & Yu, 1990). Recent synthesis work by Dai et al. (2021) further suggests that samples in the range of 200–300 are sufficient for accurate parameter estimation under typical rating-scale conditions, with larger samples providing increased robustness for model comparison. In applied IRT settings such as large-scale assessment programs, sample sizes of 500 or more are more common and support stable estimation for polytomous IRT models. With respect to DIF analyses, simulation evidence indicated that group sample sizes of approximately 200 respondents per group are generally adequate for providing adequate power for DIF detection using short scales, particularly for moderate to large DIF effects (Scott et al., 2009). Accordingly, the present pooled sample provides sufficient statistical power to support both GRM-IRT and subsequent DIF analyses.

Measures

Demographics. Across datasets included in S1, participants reported demographic information via self-report questionnaires. Demographic variables harmonized for the present analyses included age, gender, race/ethnicity, marital status, and years of education (or years in college, when available). When necessary, variables were recoded to ensure conceptual and metric equivalence across datasets before analysis.

Hostile Masculinity (HM). HM was assessed using established self-report measures that were administered across the contributing datasets. Consistent with the systematic review by Ray and Parkhill (2023), the present IDA focuses on three commonly used components of HM that were available in both datasets: the nine-item ASB scale (Burt, 1980), the seven-item SOD Scale (Nelson, 1979), and the 10-item HTW scale (Lonsway & Fitzgerald, 1995). A sample item of the SOD includes Item 4, “I have sexual relations, because like many people, I enjoy the conquest.” A sample item of HTW includes Item 1, “I feel that many times women flirt with men just to tease them or hurt them.” A sample item of ASB includes Item 6, “In a dating relationship, a woman is largely out to take advantage of a man.” Across datasets, items were rated on a 7-point Likert-type scale ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*), with higher scores indicating greater endorsement of HM beliefs. Because identical items and response formats were used across datasets, measures were directly integrated at the item level for the IDA.

Statistical Analysis

The analytic strategy for the IDA in Study 1 consists of four sequential steps. First, exploratory factor analyses (EFA) were conducted to evaluate the dimensionality of each HM subscale (i.e., SD, HTW, ASB). Second, IRT-GRMs were estimated for each subscale using the pooled sample ($N = 720$) to evaluate item functioning and estimate item parameters. Third, multiple-group GRMs were estimated as a function of study membership to formally test for DIF across studies. Fourth, when significant DIF was detected, the magnitude and source of DIF were evaluated using model fit indices, parameter differences, and inspection of item characteristic curves. Follow-up analyses then evaluated whether the detected DIF had practical implications.

Item Response Theory (IRT). IRT is a collection of latent variable models that provide information about the properties of items and the scales they comprise through the analysis of individual item responses (de Ayala, 2013). Unlike classical test theory, IRT allows item parameters to be estimated independently of sample-specific score distributions, making it particularly well suited for pooled data analysis and cross-study measurement evaluation (Embretson & Reise, 2013). Within IRT, the association between a latent trait and item responses is typically formalized using a logistic function, allowing endorsement probabilities to vary smoothly across the latent continuum. In dichotomous IRT models such as the two-parameter logistic (2PL) model, this formulation produces item discrimination parameters that index how strongly items relate to the trait and item difficulty parameters that locate items along the trait continuum (de Ayala, 2013).

The graded response model (GRM; Samejima, 1969) extends this logistic framework to ordered polytomous response data. Similar to the 2PL model, the GRM estimates a single discrimination parameter per item, reflecting the strength of association between the item and the underlying latent trait. Whereas the 2PL model specifies a single difficulty parameter for each dichotomous item, the GRM accommodates multiple ordered response categories (e.g., Likert-scale items such as "*Strongly Disagree*," "*Disagree*," "*Neutral*," "*Agree*," and "*Strongly Agree*") by estimating a set of threshold parameters that define category boundaries along the latent trait continuum (Ayala, 2013; Dai et al., 2021; Nering & Ostini, 2011). For an item with *K* ordered response categories, the GRM specifies *K-1* threshold parameters, each corresponding to the transition point at which respondents are more likely to endorse a higher response category rather than a lower one. These threshold parameters are modeled cumulatively, such that the probability of responding at or above a given category is expressed as a logistic function of the latent trait, resulting in a monotonic ordering of category boundaries by design (Penfield, 2014).

To address Objective 1, GRM-IRT models were estimated for each HM subscale to evaluate item-level measurement structure. The graded response model (GRM; Samejima, 1969) was selected because it is designed for ordered polytomous response data and is appropriate for HM items measured on a 7-point Likert-type scale (1 = *Strongly Disagree* to 7 = *Strongly Agree*). For each HM subscale, alternative GRM specifications were compared to determine the most appropriate dimensional structure. Specifically, a unidimensional GRM and a multidimensional GRM were estimated and evaluated to assess whether item responses are best represented by a single latent dimension or by multiple latent dimensions.

All IRT analyses were conducted in R (Version 4.3.2; R Core Team, 2023) using RStudio as the development environment. Preliminary dimensionality checks and factor-analytic diagnostics for ordinal indicators were conducted using the psych package

(Revelle, 2026). GRMs were estimated using the *mirt* package (Chalmers, 2012). Model fit was evaluated using M_2 statistics (Maydeu-Olivares & Joe, 2006), which is conceptually similar to the χ^2 test of overall model fit in SEM (Zein & Akhtar, 2025). Additional goodness-of-fit indices included the root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and standardized root mean square residual (SRMR). Following the guidelines from Schermelleh-Engel et al. (2003), values of RMSEA $\leq .05$ and SRMR $< .05$, along with CFI and TLI $\geq .97$, indicate good model fit, whereas RMSEA between .05 and .08, SRMR between .05 and .10, and CFI and TLI between .95 and .97 indicate acceptable fit. Model comparisons were based on standard IRT fit indices, including the −2 log-likelihood (−2LL), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and sample-size-adjusted BIC (SABIC), with lower values indicating better fit (de Ayala, 2013; Raftery, 1995). When appropriate, likelihood ratio tests (LRTs) were used for comparisons between nested models. Item-level fit was evaluated using $S\text{-}X^2$ item fit statistics (Orlando & Thissen, 2000) and inspection of item characteristic curves (ICCs). Item functioning was examined by inspecting item discrimination parameters and evaluating whether threshold parameters show an orderly progression across response categories. In the GRM framework, discrimination parameters (α) greater than 1.0 indicate strong item discrimination, whereas threshold (difficulty) parameters are typically expected to fall within the range of −3 to 3 (Rezapour et al., 2021). Items exhibiting low discrimination, disordered thresholds, or evidence of poorly functioning response categories were flagged for further evaluation or potential removal prior to final model specification.

Differential Item Functioning. DIF analysis refers to a form of item-level measurement invariance testing that examines whether individual items function equivalently across groups, controlling for individuals' location on the underlying latent trait (Lord, 1980). An item exhibits DIF when respondents from different groups with equivalent levels of the latent construct have different probabilities of endorsing a given response category. Within an IRT framework, Likelihood-Ratio (LR) DIF testing evaluates item invariance by comparing nested multiple-group models in which item parameters are constrained to equality across groups versus models in which those parameters are allowed to vary. Significant improvements in model fit following release of item-specific constraints are interpreted as statistical evidence of DIF (Cohen et al., 1996; Teresi & Fleishman, 2007).

DIF assessment is particularly important in IDA, where item-level data are pooled across independent studies that may differ in sampling frames, recruitment contexts and measurement implementation. Even when identical items are administered, between-study heterogeneity can lead to differences in how items function, indicating a lack of invariance in how the latent construct is measured across studies. Failure to account for DIF may lead to biased parameter estimates, distorted latent trait scores, and spurious conclusions when datasets are pooled (Millsap, 2011).

In the current study, DIF analyses were conducted in R (Version 4.3.2; R Core Team, 2023) using RStudio, with the *mirt* package (Chalmers, 2012). These analyses were conducted following the estimation of GRM IRT models for each HM subscale. After establishing an acceptable baseline GRM model for each subscale, DIF were evaluated using an iterative multiple-group IRT-LR approach, with study membership serving as the grouping variable. An initial multiple-group model constrained all item parameters to equality across studies to establish a provisional anchor set. Items showing the strongest evidence of DIF were then freed sequentially, with the model re-estimated at each step, until no additional items showed evidence of DIF (Woods, 2008). Items showing statistically significant DIF underwent substantive evaluation to determine whether the DIF reflects study-specific methodological differences, population differences, or construct-relevant variability. When DIF was detected, its practical and interpretive importance was evaluated using EAP score concordance and substantive item review. DIF results were used as diagnostic evidence rather than as the sole basis for item removal. The retained item pool identified through the full set of item-level analyses served as the basis for subsequent latent structure analyses.

Results

The pooled Study 1 sample included 720 men drawn from two independent datasets. Project Become 1 contributed 229 participants recruited from a psychology subject pool at a large Midwestern university, whereas Project Become 2 contributed 491 participants recruited online through MTurk. Consistent with these recruitment differences, Become 1 consisted primarily of traditional college-aged undergraduate students, whereas Become 2 included a broader adult sample with greater variability in age, education, income, and geographic region. Given these differences in sampling frame and participant characteristics, Study 1 evaluated the item-level functioning of the HM measures before using the pooled data in subsequent latent structure analyses.

Results are presented separately for HTW, ASB, and SOD. For each measure, dimensionality evidence was presented first, followed by GRM-IRT results and DIF analyses comparing item functioning across the two datasets.

Measurement Analysis

Hostility Towards Women (HTW)

Dimensionality Check with Factor Analysis. Prior to conducting the factor analysis, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were performed to evaluate the suitability of the data for factor analysis. Results of KMO indicated excellent adequacy (KMO = .90), and Bartlett's test of sphericity was significant, $\chi^2(45) = 4742.06$, $p < .001$, suggesting that the correlation matrix was appropriate for factor analysis. To determine the number of factors underlying the HTW items, eigenvalues and a scree plot of the correlation were first examined. The eigenvalues and scree plot (see Figure 1) indicated that there were two factors that exceeded the Kaiser criterion of 1.0 ($\lambda_1 = 5.05$, $\lambda_2 = 1.40$), together accounting for 64.5% of the total variance. Thus, the following EFA assessments were conducted using maximum likelihood estimation with both one-factor and two-factor solutions. The one-factor model showed poor fit (RMSEA = .152, 90% CI [.141, .163], TLI = .79, BIC = 26.35, SRMR = .09), whereas the two-factor model demonstrated substantially improved fit (RMSEA = .097, TLI = .915, BIC = 26.35). In the one-factor solution, most items loaded strongly on the general factor, with two items (Item 2 and Item 3) showing relatively weak loadings. The two-factor solution also revealed a clear and interpretable pattern, with Item 1 to Item 6 loading primarily on the first factor and Item 7 to Item 10 loading on the second factor (see Table 3). The two factors were moderately correlated (r = .64), indicating that they are related but distinguishable.

Figure 1

Scree plot for the Hostility Toward Women Items

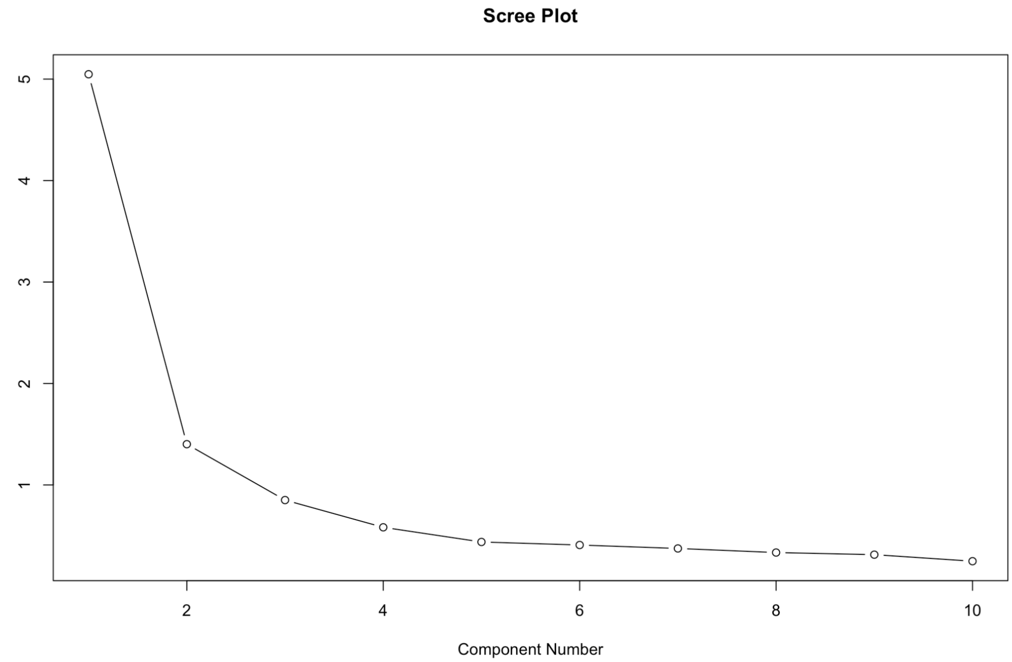


Table 3

Standardized Factor Loadings for the Two-Factor HTW Solution

Item	Factor 1	Factor 2
HTW1	0.42	0.32
HTW2	0.76	--
HTW3	0.45	--
HTW4	0.71	--
HTW5	0.68	--
HTW6	0.7	--
HTW7	--	0.75
HTW8	--	0.75
HTW9	--	0.85
HTW10	--	0.88

Note. Values are standardized pattern coefficients from maximum likelihood exploratory factor analysis. Loadings below |.30| are omitted. HTW = hostility toward women. The two factors were correlated, r = .68.

IRT Results. Based on the factor analysis results, a multidimensional GRM was estimated using the factor structure identified in the previous EFA (i.e., Items 1 to Item 6 loading on Factor 1 and Items 7 to Item 10 loading on Factor 2. The model, however, did not fit the data well, $M_2(34) = 252.249$, $p<.001$, RMSEA = .096, SRMR = .086, TLI = .957, CFI = .967. Specifically, the RMSEA was above the .08 cutoff, suggesting some misfit, while CFI and TLI were both above .95, indicating good fit. The SRMR, which reflects the average effect size of model misfit (Maydeu-Olivares, 2015), was also above the conventional .05 cutoff. These indices suggested the presence of model misfit, but the magnitude of the misfit was not severe. Moreover, when compared with the one-dimensional model, $M_2(35) = 650.97$, $p < .001$, RMSEA = .159, SRMR = .102, CFI = .907, TLI = .881, the two-dimensional solution demonstrated substantially better fit, $\Delta\chi^2(1) = 288.91$, $p < .001$, ΔAIC = 286.91, ΔBIC = 282.37. Therefore, instead of rejecting the two-dimensional model, subsequent analyses were conducted to identify the potential sources of misfit. Item-fit statistics based on the $S\text{-}X^2$ statistic (Orlando & Thissen, 2000) were then examined to further evaluate model fit at the item level. In the $S\text{-}X^2$ statistic, a nonsignificant result is typically expected, which indicates that the observed response pattern does not significantly deviate from those predicted by the model (Zein & Akhtar, 2025). Results showed that all items exhibited good item-level RMSEA values (≤ .05), suggesting that each item individually fits the underlying IRT model well. Several items (Items 1, 2, 3, 4, and 9) showed significant $S\text{-}X^2$, which is common in large samples, where the χ^2 test becomes overly sensitive to minor deviations from the model (Kline, 2023). Next, local dependence was evaluated using Yen's Q_3

statistic (Yen, 1984), which is considered more sensitive for detecting local dependence of polytomous data than the LD χ^2 statistic (Chen & Thissen, 1997; Zein & Akhtar, 2025). The results indicated that the residual correlation for one item pair (Item 2 and 3: r = .41) exceeded the commonly used cutoff of 0.20, suggesting that local dependence may have partially contributed to the observed model misfit.

Given that local dependence was generally limited and multidimensionality had already been evaluated, additional potential sources of model misfit were examined, including sparse response categories and extreme discrimination parameters. Inspection of category endorsement frequencies revealed that several upper response categories were rarely endorsed across items (e.g., for Items 2, 3 and 7, the highest response category was endorsed by no more than 2% of respondents). However, sparse category usage did not appear to be a major concern, because all response options were endorsed by more than 1% of respondents. In addition, several items (i.e., Items 5, 6, 7, 9, and 10) demonstrated large discrimination parameters (e.g., $a > 3.0$), indicating steep item characteristic curves and strong differentiation along the latent trait. Item parameter estimates are presented in Table 4. Such values, though, are not surprising, as the HTW items involve strong statements reflecting extreme hostile attitudes that are likely to be endorsed by only a small subset of respondents. Given that other diagnostics (i.e., local dependence and response category functioning) also did not reveal substantial issues, the two-factor GRM model and all items were retained at this stage.

Table 4

Item Parameter Estimates for the Two-Dimensional GRM of the HTW Items

Item	Factor	a	b_1	b_2	b_3	b_4	b_5	b_6
HTW1	F1	1.66	-1.7	-0.56	-0.06	0.6	1.44	2.46
HTW2	F1	1.07	-2.65	-0.61	0.31	1.41	2.34	4.04
HTW3	F1	0.61	-5.17	-1.65	0.17	3	5.09	7.76
HTW4	F1	2.63	-1.4	-0.44	0.06	0.72	1.42	2.21
HTW5	F1	3.82	-0.78	0.06	0.41	0.95	1.48	2.06
HTW6	F1	3.74	-0.68	0.04	0.44	0.95	1.5	1.99
HTW7	F2	3.18	-0.23	0.48	0.82	1.28	1.62	2.39
HTW8	F2	2.77	-0.38	0.26	0.5	1.25	1.61	2.09
HTW9	F2	3.24	-0.03	0.59	0.89	1.16	1.69	2.18
HTW10	F2	3.15	0.18	0.75	0.99	1.36	1.73	2.25

Note. GRM = graded response model; HTW = hostility toward women. Items 1-6 loaded on the HTW distrust factor, and Items 7-10 loaded on the HTW resentment factor. a = item discrimination parameter. b_1-b_6 = threshold parameters. Higher a values indicate stronger item discrimination.

Item functioning was then further examined by examining the category characteristic curves (CCCs; See Figure 2) and item information functions (See Figure 3). Overall, the CCCs followed the expected ordered pattern, with adjacent categories being the most probable across increasing levels of the latent trait. The only exception was Items 2 and 3. Item 2 demonstrated steep threshold transitions and abrupt changes in response probabilities across the latent trait continuum. Such patterns suggest highly sensitive category boundaries and potential instability in threshold estimation. In contrast, Item 3 showed relatively flat and overlapping response probability regions, which indicates weaker category differentiation across levels of the latent trait. This pattern was consistent with its low discrimination parameter and low communality estimate, suggesting that the item contributed relatively limited information to the latent construct. These two items were therefore flagged for further inspection.

Figure 2

Category Characteristic Surfaces for the HTW Items

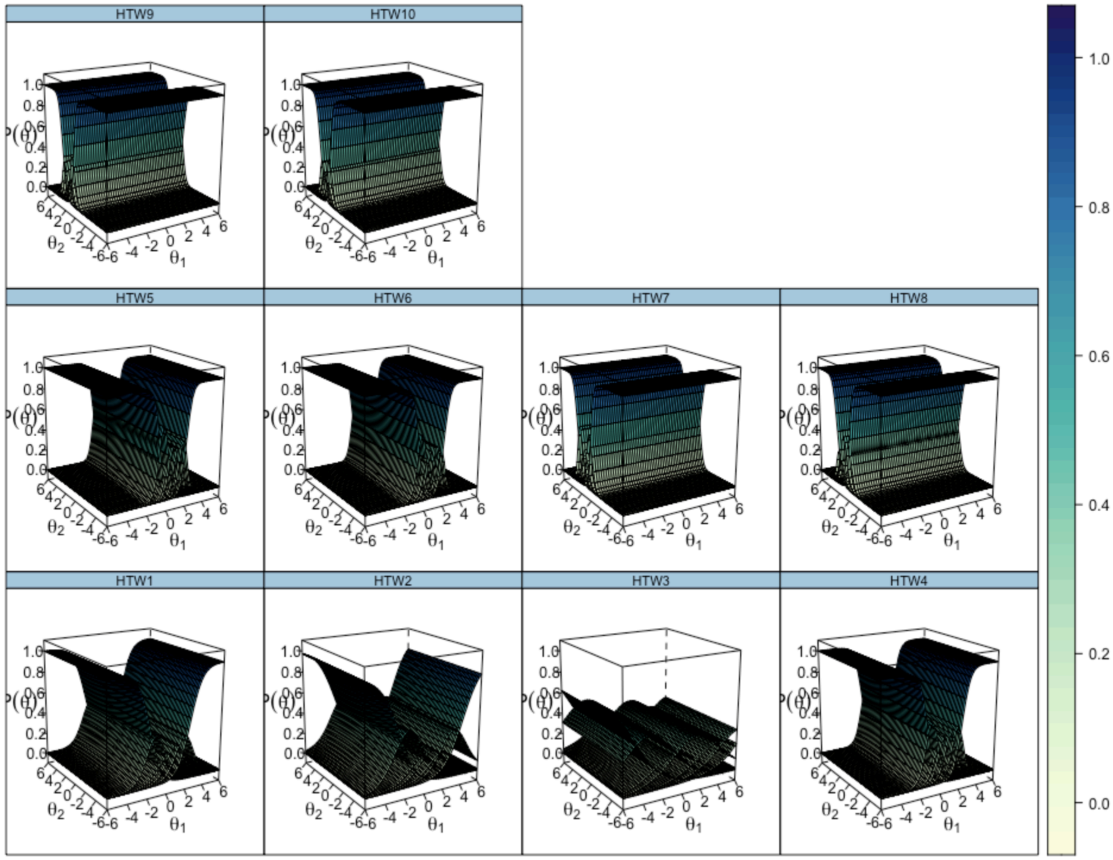
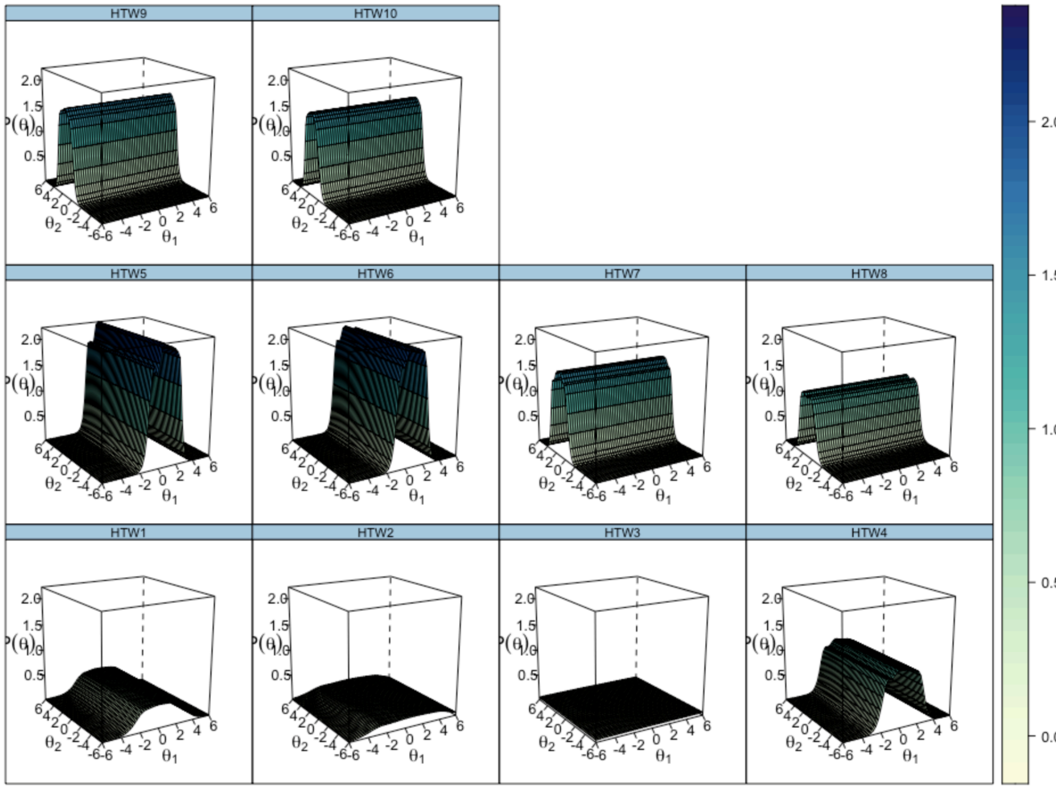


Figure 3

Item Information Surfaces for the HTW Items



DIF Results. To examine potential DIF between two sample groups (Project Become 1 & 2), a series of multi-group GRM models were estimated. First, a baseline configural model was specified in which item discrimination and threshold parameters were freely estimated across groups. A second model constrained item slopes and thresholds to be equal across groups while allowing latent means and variances to vary freely, representing a model of full measurement invariance. A likelihood ratio test comparing the configural and fully invariant models was then conducted to evaluate whether constraining item parameters across groups resulted in a significant reduction in model fit. Results indicated that the fully invariant model was significantly worse than the configural model ($\Delta\chi^2(52) = 158.95$, $p <.001$), suggesting potential DIF across groups. Following the recommendation of early IDA research (Curren et al., 2018), the sensitivity of expectation a posteriori (EAP) was then tested to evaluate the practical impact of DIF affecting latent factor score estimates. Comparisons of EAP scores derived from the constrained and unconstrained models demonstrated a high degree of concordance, suggesting that the practical impact of the statistically significant DIF was negligible. Subsequent item-level DIF analyses revealed that only HTW2 demonstrated significant item-level DIF, $\chi^2(6) = 46.87$, $p < .001$. In a follow-up purification check, HTW2 was temporarily excluded from the DIF model; after this diagnostic step, no remaining items demonstrated significant DIF. Overall, the detected noninvariance appeared localized rather than systematic across the full HTW measure, and EAP score concordance suggested that its practical impact was negligible.

Adversarial Sexual Beliefs (ASB)

Dimensionality Check with Factor Analysis. The KMO measure of sampling adequacy indicated excellent adequacy (KMO = .92), and Bartlett's test of sphericity was significant, $\chi^2(36) = 3495.50$, $p < .001$, suggesting that the correlation matrix was appropriate for factor analysis. The eigenvalues indicated a dominant first factor, with all subsequent eigenvalues below 1.00. Visual inspection of the scree plot also supported this unidimensional structure (see Figure 4). Based on this preliminary assessment, an EFA was conducted using maximum likelihood estimation. Results of EFA also supported a single-factor structure, with all items loaded positively on the factor and standardized loadings ranging from .43 to .82 (see Table 5). The one-factor model accounted for 53% of the total variance, with acceptable model fit, RMSEA = .106, 90% CI [.094, .119], TLI = .918, and SRMR = .04.

Figure 4
Scree plot for the Adversarial Sexual Beliefs Items

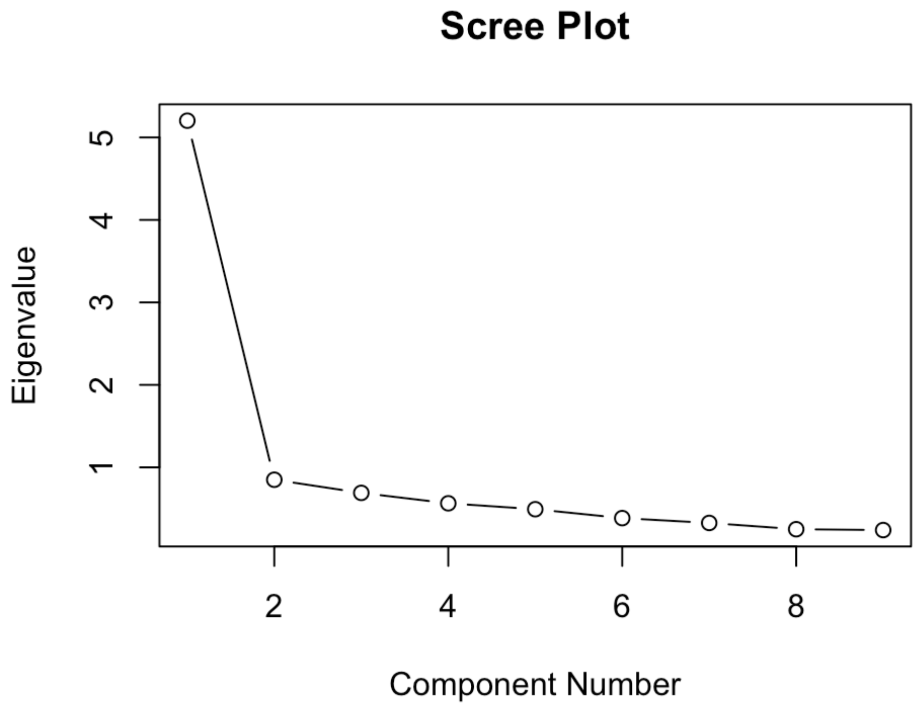


Table 5

Standardized Factor Loadings for the One-Factor ASB Solution

Item	ASB
ASB1	0.78
ASB2	0.71
ASB3	0.82
ASB4	0.76
ASB5	0.43
ASB6	0.82
ASB7	0.55
ASB8	0.82
ASB9	0.77

Note. Values are standardized factor loadings from the one-factor exploratory factor analysis. ASB = adversarial sexual beliefs.

IRT Results. Given the results from factor analysis, the ASB items were tested with a one-dimensional GRM model. The model, however, did not fit the data well, $M_2(27) = 224.104$, $p < .001$, RMSEA = .101, SRMR = .068, TLI = .962, CFI = .971. Specifically, the RMSEA suggested poor fit, while CFI and TLI indicated good fit. The SRMR was slightly above the conventional .05 cutoff (Maydeu-Olivares, 2015). Together, these indices suggested the presence of model misfit, but the magnitude of the misfit was not severe. Therefore, instead of rejecting the model, subsequent analyses were conducted to identify the potential sources of misfit. Item-fit statistics based on the $S\text{-}X^2$ statistic (Orlando & Thissen, 2000) were then examined to further evaluate model fit at the item level. In the $S\text{-}X^2$ statistic, a nonsignificant result is typically expected, which indicates that the observed response pattern does not significantly deviate from those predicted by the model (Zein & Akhtar, 2025). Results showed that all items exhibited good item-level RMSEA values (≤ .05), suggesting that each item individually fits the underlying IRT model well. However, several items (Items 1, 4, 5, 7, and 9) showed significant $S\text{-}X^2$, which is common in large samples, where the χ^2 test becomes overly sensitive to minor deviations from the model (Kline, 2023). Next, local dependence was evaluated using Yen's Q_3 statistic (Yen, 1984), which is considered more sensitive for detecting local dependence of polytomous data than the LD χ^2 statistic (Chen & Thissen, 1997; Zein & Akhtar, 2025). The results indicated that none of the residual correlations exceeded the commonly used cutoff of 0.2, suggesting that local dependence was not responsible for the observed model misfit.

Given that local dependence was not detected, other potential sources of misfit were examined, including multidimensionality, sparse response categories, and extreme discrimination parameters. Although dimensionality was checked earlier with factor analysis, an additional exploratory GRM with two factors was estimated to further evaluate potential multidimensionality. The two-factor solution demonstrated good fit, $M_2(19)=63.090$, $p<.001$, RMSEA=.057, SRMR=.055, TLI=.988, CFI=.994, which substantially improved overall model fit relative to the one-factor model (e.g., RMSEA decreased from .101 to .057, CFI increased from .971 to .994). However, the information criteria increased for the two-factor solution (ΔAIC = 139.62, ΔBIC = 103.26), suggesting that the improvement in fit did not justify the added model complexity. In addition, inspection of the item parameters indicated that all items continued to load strongly on the primary factor, whereas loadings on the second factor were weaker and less consistent across items. The improved model fit likely reflected minor residual covariance among items rather than the presence of two substantively distinct latent dimensions. The unidimensional solution was therefore retained. Meanwhile, this evidence of residual covariance further supported the use of ESEM in later analyses, which allowed item cross-loadings and provided greater flexibility in modeling complex latent structures without forcing items to load on only a single factor. Inspection of the response category distributions indicated that several upper response categories were infrequently endorsed (i.e., by less than 5% of respondents), particularly for the most extreme response options. Although all response categories were endorsed by more than 1% of respondents, sparse endorsement of upper categories may have contributed to some instability in parameter estimation and overall model misfit. In addition, three items (Items 3, 6, and 8) showed very high discrimination parameters (i.e., $a \geq 3.0$), indicating steep item characteristic curves and strong differentiation along the latent trait. Table 6 presents item parameter estimates of ASB items. Such values, though, are not surprising, as the ASB items involve strong statements reflecting extreme hostile attitudes that are likely to be endorsed by only a small subset of respondents. Given that other diagnostics (e.g., local dependence and response category functioning) also did not reveal substantial issues, the one-factor GRM model and all items were retained at this stage.

Table 6
Item Parameter Estimates for the One-Dimensional GRM of the ASB Items

Item	a	d_1	d_2	d_3	d_4	d_5	d_6
ASB1	2.88	1.05	-1.2	-2.21	-3.31	-4.72	-6.01
ASB2	2.28	1.37	-0.95	-1.86	-3.04	-4.14	-5.4
ASB3	3.48	0.73	-1.76	-2.94	-4.37	-5.9	-7.91
ASB4	2.6	1.98	-0.06	-1.22	-2.2	-3.7	-5.55
ASB5	0.89	2.63	1.57	1.13	-0.11	-1.34	-2.77
ASB6	3.52	1.51	-1.33	-2.88	-4.32	-5.74	-7.33
ASB7	1.41	0.96	-0.41	-1.18	-1.9	-3.39	-5.05
ASB8	3.37	2.08	-0.31	-1.48	-2.71	-4.58	-6.58
ASB9	2.82	2.21	-0.15	-1.06	-2.33	-3.93	-5.99

Note. GRM = graded response model; ASB = adversarial sexual beliefs. a = item discrimination parameter. d_1-d_6 = threshold parameters. Higher a values indicate stronger item discrimination.

Further inspection of the threshold surfaces and item probability functions (See Figure 5) suggested that overall, the CCCs followed the expected ordered pattern, with adjacent categories being the most probable across increasing levels of the latent trait. The only exception was Item 5, whose curves were flatter and wider, with greater overlap among the middle categories and less distinct separation between adjacent response options. This pattern is consistent with the item's low discrimination parameter estimates (a = 0.89). Inspection of the item information functions indicated that most items provided the greatest measurement precision at moderate to moderately high levels of latent trait (See Figure 6). Items 3, 6, and 8 contributed the most information, consistent with their high discrimination parameters. In contrast, Items 5 and 7 contributed comparatively little information, as reflected in their relatively flat item information curves. These two items were therefore flagged for further evaluation.

Figure 5

Category Characteristic Surfaces for the ASB Items

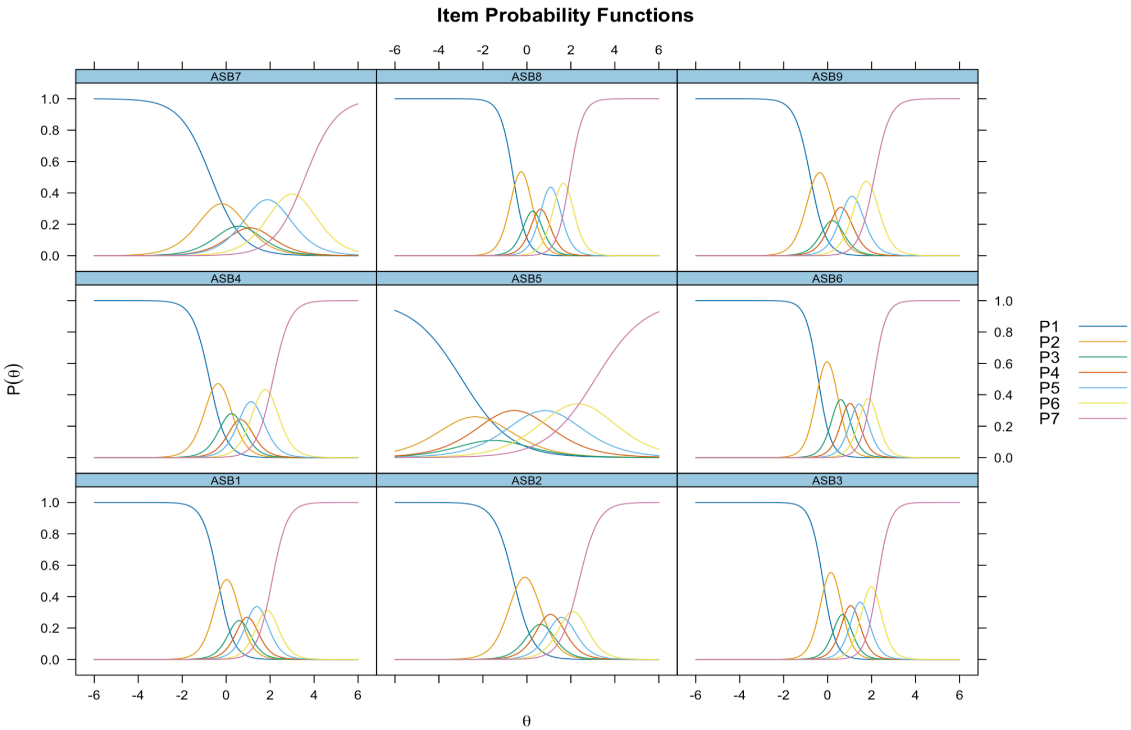
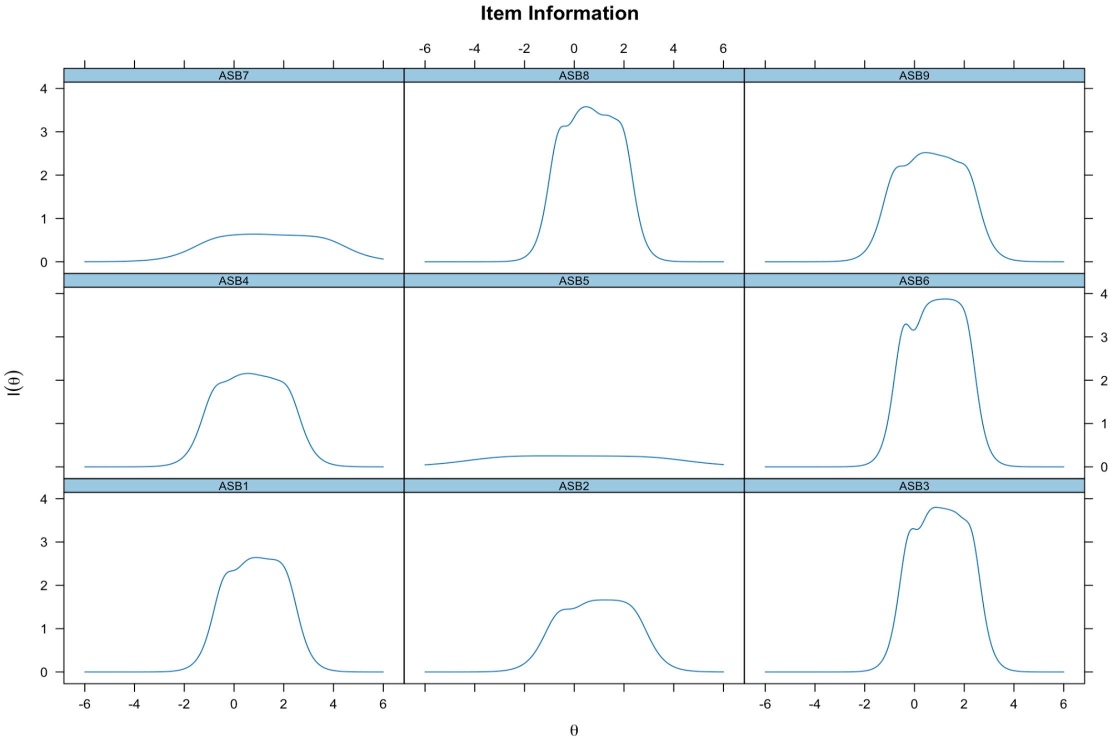


Figure 6

Item Information Curves for the ASB Items



DIF Results. To examine potential DIF between two sample groups (Project BECOME 1 & 2), a series of multi-group GRM models were estimated. First, a baseline configural model was specified in which item discrimination and threshold parameters were freely estimated across groups. A second model constrained item slopes and thresholds to be equal across groups while allowing group means and variances to be freely estimated, representing a model of full measurement invariance. A likelihood ratio test comparing the configural and fully invariant models indicated that the fully invariant model did not fit significantly worse than the configural model ($\Delta\chi^2(52) = 68.65$, $p = .061$), suggesting that no substantial DIF was found across groups. As suggested by early research using IDA (Curren et al., 2018), the sensitivity of expectation a posteriori (EAP) was tested to evaluate the practical impact of DIF affecting latent factor score estimations. EAP ASB scores were derived and then compared with unadjusted raw ASB total scores. There was a high degree of concordance between EAP and unadjusted ASB scores, further supporting that the impact of statistically significant DIF was negligible. Although the overall model comparison did not indicate substantial scale-level DIF, item-level follow-up analyses identified localized evidence of DIF for Items 2, 9, 1, 3, and 4. Stepwise purification was then conducted as a diagnostic follow-up. Item 2 was temporarily excluded first ($\chi^2(7) = 17.89$, $p < .001$), followed by Item 9 ($\chi^2(7) = 24.59$, $p < .001$) and Item 1($\chi^2(7) = 20.23$, $p < .001$), to evaluate whether localized DIF remained after accounting for the strongest DIF-flagged items. After these diagnostic exclusions, no items showed significant DIF. The localized item-level DIF appeared to be driven by a small subset of ASB items rather than reflecting pervasive noninvariance across the scale.

These findings suggest that the limited DIF observed at the item level did not reflect substantial noninvariance across the overall ASB scale.

Sexual Dominance (SOD)

Dimensionality Check with Factor Analysis. Results of the KMO measure of sampling adequacy indicated excellent adequacy, and Bartlett's test of sphericity was significant, suggesting that the correlation matrix was appropriate for factor analysis. The inspection of eigenvalues and scree plot (see Figure 7) supported the evaluation of a potential three-factor structure, with three factors exceeding the Kaiser criterion of 1.0 ($\lambda_1 = 5.03$, $\lambda_2 = 3.77$, $\lambda_3 = 1.02$), together accounting for 61.3% of the total variance. Therefore, one-factor, two-factor, and three-factor EFAs were subsequently evaluated. Consistent with the eigenvalue and scree plot results, the one-factor solution demonstrated poor fit (RMSEA = .193, 90% CI [.187, .199], TLI = .387, BIC = 2097.59, SRMR = .22), supporting the multidimensional structure of SOD items. In contrast, the two-factor solution demonstrated substantially improved fit (RMSEA = .090, 90% CI [.083, .097], TLI = .868, BIC = -0.13, SRMR = .04) and yielded a relatively simple and interpretable structure. Specifically, Items 1, 3, 6, 8, 10, 12, 14, and 16 loaded primarily on the first factor, whereas Items 2, 4, 5, 7, 9, 11, and 15 loaded primarily on the second factor. Item 13 demonstrated a modest cross-loading across both factors. The two factors were only weakly correlated ($r = .09$), suggesting that they represent two distinct dimensions. Although the three-factor solution demonstrated better overall fit (RMSEA = .075, 90% CI [.068, .083], TLI = .906, BIC = -120.29, SRMR = .03), inspection of the factor loading pattern suggested that the additional factor primarily split the first factor from the two-factor solution into two correlated components ($r = .69$). In addition, two more items (Item 6 and 14) demonstrated increased cross-loadings in the three-factor solution, resulting in reduced interpretability and less distinct factor separation. Therefore, the two-factor solution was retained as the most parsimonious and theoretically interpretable representation of the SOD items. One of the factors, consisting of Items 2, 4, 5, 7, 9, 11, and 15, along with the cross-loaded Item 13, captured the subset of items traditionally used to operationalize hostile masculinity in prior Confluence Model research (Smith et al., 2015). Given that the primary focus of the current study was hostile masculinity measurement, subsequent analyses retained only this eight-item subset of SOD items for further psychometric evaluation. Standardized factor loadings for the retained eight-item SOD subset are presented in Table 8.

Figure 7

Scree plot for the Sexual Dominance Items

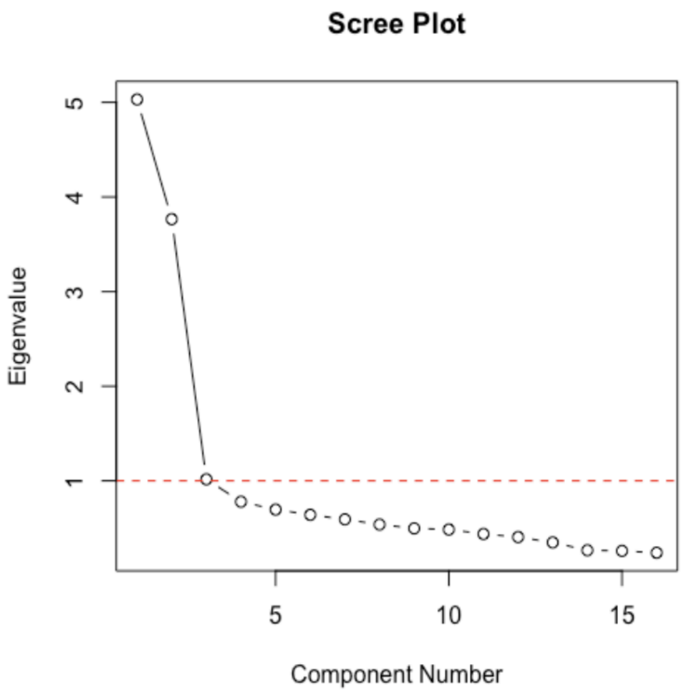


Table 7
Standardized Factor Loadings for the Two-Factor SOD Solution

Item	Factor 1	Factor 2
SOD12	0.82	--
SOD10	0.78	--
SOD14	0.74	--
SOD16	0.69	--
SOD6	0.67	--
SOD1	0.62	--
SOD3	0.61	--
SOD8	0.5	--
SOD13	0.46	0.32
SOD5	--	0.81
SOD7	--	0.8
SOD15	--	0.78
SOD11	--	0.71
SOD4	--	0.67
SOD9	--	0.6
SOD2	--	0.58

Note. Values are standardized pattern coefficients from exploratory factor analysis using a correlation matrix. Loadings below |.30| are omitted. SOD = sexual dominance

Table 8

Standardized Factor Loadings for the Retained Eight-Item SOD Subset

Item	SOD
SOD2	0.59
SOD4	0.65
SOD5	0.8
SOD7	0.81
SOD9	0.6
SOD11	0.71
SOD13	0.37
SOD15	0.8

Note. Values are standardized pattern coefficients from exploratory factor analysis. Loadings below |.30| are omitted. SOD = sexual dominance

IRT Results. A unidimensional GRM was subsequently fitted to this subset of SOD items (Items 2, 4, 5, 7, 9, 11, 13, and 15). The model showed evidence of misfit $M_2(20) = 294.45$, $p < .001$, RMSEA = .141, SRMR = .072, TLI = .906, CFI = .933, particularly according to RMSEA, although item-level diagnostics suggested the misfit was not concentrated in severe item-level problems. The SRMR was also slightly above the conventional .05 cutoff (Maydeu-Olivares, 2015). These results suggested the presence of some degree of model misfit. Since the overall magnitude of the misfit was not severe, subsequent analyses were conducted to further evaluate potential sources of misfit. Results from the item-level $S\text{-}X^2$ statistics indicated that all items demonstrated good item fit with relatively low item-level RMSEA values ($\leq .028$). Although three items (Items 2, 7, and 11) showed significant $S\text{-}X^2$, the corresponding RMSEA.$S\text{-}X^2$ values remained small (i.e., $\leq .05$), suggesting that the magnitude of item-level misfit was minimal. Thus, these significant $S\text{-}X^2$ statistics may reflect the sensitivity of chi-square-based fit indices to large sample sizes rather than substantial model misfit. Next, local dependence was evaluated using Yen's Q3 statistic. Results indicated that most residual correlations fell below the commonly used .20 cutoff, suggesting minimal evidence of local dependence. Although one residual correlation slightly exceeded this threshold (Items 4 and 5: $r = .237$), the overall pattern of residual correlations did not suggest substantial local dependence among the retained SOD items.

Although the dimensionality was checked earlier with the whole SOD scale, additional exploratory GRM with two factors was estimated to further evaluate potential multidimensionality. The two-factor solution demonstrated improved overall model fit,

$M_2(13) = 63.140$, $p < .001$, RMSEA = .075, SRMR = .043, TLI = .974, CFI = .988, as compared to the one-factor model. However, all except two items (i.e., Items 4 and 5) continued to load strongly on one primary factor, suggesting that multidimensionality was unlikely to be the primary source of the observed model misfit. The misfit might, instead, reflect minor residual covariance among items. Inspection of item category distributions and discrimination parameters suggested several potential contributors to the observed misfit. Specifically, several upper categories were endorsed by less than 5% of respondents, particularly for the most extreme response options (e.g., Category 7 endorsement ranged from 3.5% to 8.5%). In addition, discrimination parameters ranged from 0.74 to 3.08, with three items (i.e., Items 5, 11, and 15) demonstrating high discrimination parameters ($a \geq 2.0$) and one item (i.e., Item 7) demonstrating exceptionally high discrimination ($a \geq 3.0$). Item parameter estimates are presented in Table 9. Nevertheless, because all response categories were endorsed by more than 3% of respondents, item-level fit statistics generally remained acceptable, and local dependence diagnostics did not reveal substantial violations of model assumptions, the one-factor GRM model and all retained SOD items were retained at this stage.

Table 9

Item Parameter Estimates for the One-Dimensional GRM of the Retained SOD Items

Item	a	b_1	b_2	b_3	b_4	b_5	b_6
SOD2	1.45	1.83	0.97	0.46	-0.54	-1.81	-3.18
SOD4	1.7	1.67	0.83	0.06	-0.94	-2.26	-3.63
SOD5	2.87	1.81	0.5	-0.64	-2.16	-3.72	-6.07
SOD7	3.08	1.82	0.37	-0.75	-1.92	-3.46	-5.21
SOD9	1.62	0.75	-0.23	-0.84	-1.98	-2.84	-3.97
SOD11	2.19	1.72	0.38	-0.45	-1.75	-3.11	-4.51
SOD13	0.74	2.27	1.66	1.15	0.38	-0.48	-1.58
SOD15	2.88	1.55	0.2	-0.87	-2.06	-3.3	-4.54

Note. GRM = graded response model; SOD = Sexual Dominance. a = item discrimination parameter. b_1-b_6 = threshold parameters. Higher a values indicate stronger item discrimination.

According to the CCCs (see Figure 8), response categories were generally ordered appropriately across increasing levels of the latent trait continuum, suggesting adequate threshold functioning for the retained SOD items. Inspection of the item information curves further indicated that all retained SOD items provided information across the latent HM continuum (see Figure 9). Specifically, Items 5, 7, and 15 demonstrated relatively narrow and overlapping middle response categories along with sharply peaked information functions consistent with their high discrimination parameters. In contrast, Items 2, 4, and 9 demonstrated broader and relatively flatter information curves with less distinct middle category differentiation, suggesting lower precision and weaker differentiation across trait levels. Item 13 demonstrated the broadest and flattest information curve, consistent with its comparatively lower discrimination parameter and earlier cross-loading pattern observed in the factor analyses, and was thus flagged for further evaluation.

Figure 8

Category Characteristic Surfaces for the SOD Items

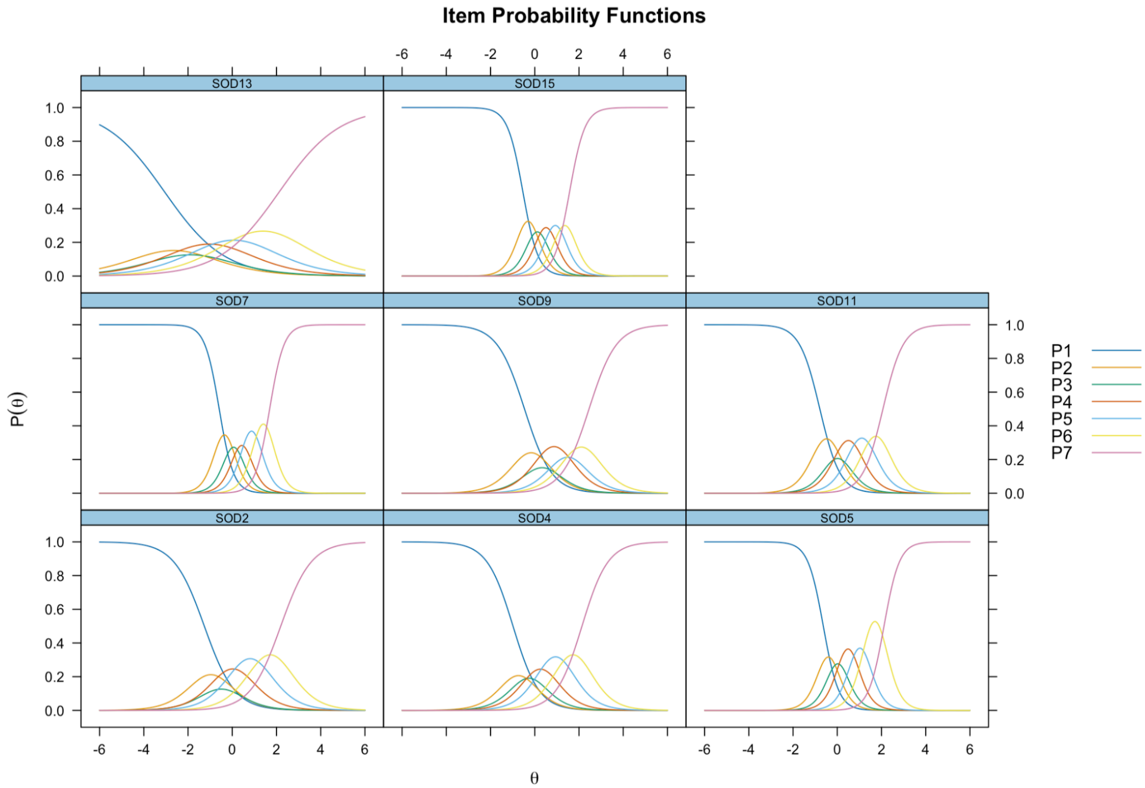
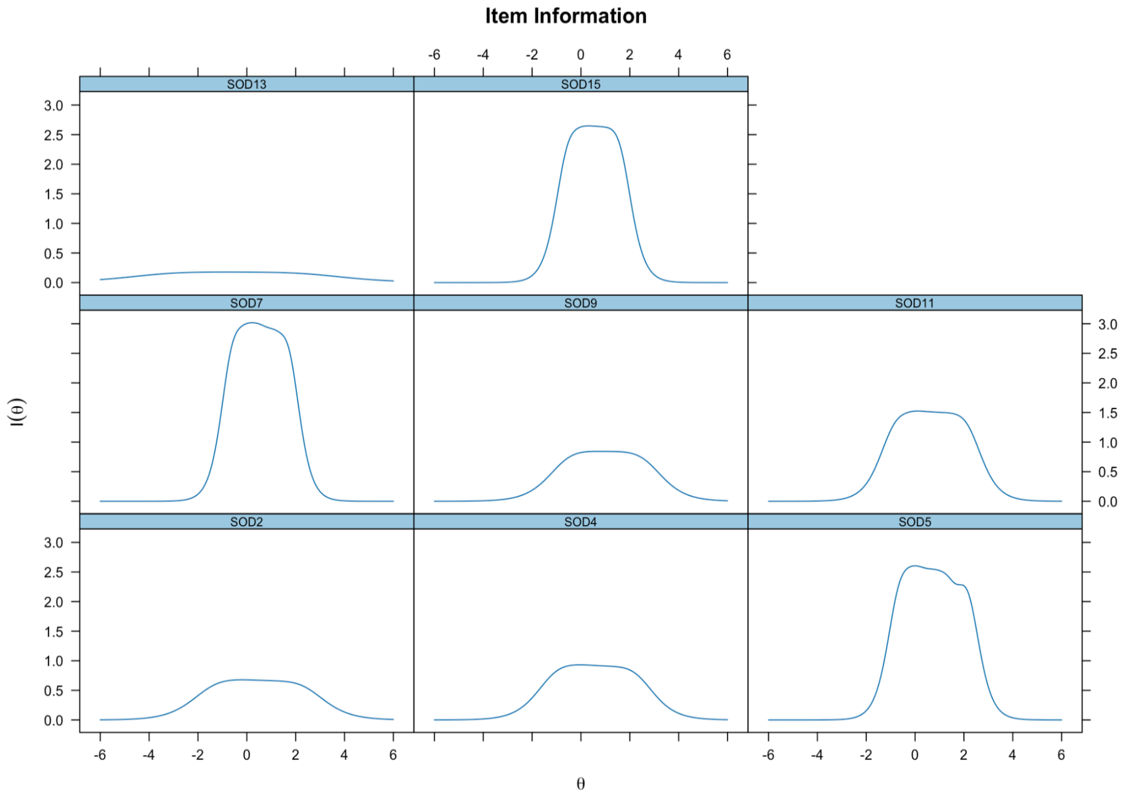


Figure 9

Item Information Curves for the SOD Items



DIF Results. The likelihood ratio test comparing the configural and fully invariant models indicated that the fully invariant model fit significantly worse than the configural model ($\Delta\chi^2(48) = 43.31$, $p < .001$), suggesting potential DIF across groups. Subsequent item-level DIF analyses revealed that Items 2, 4, 5 and 13 demonstrated significant DIF. However, comparisons between EAP scores derived from the constrained and unconstrained models demonstrated a very high degree of concordance (r = .998), suggesting that the practical impact of DIF on latent score estimation was negligible. Subsequent item-level DIF analyses revealed that Items 2, 4, 5 and 13 demonstrated significant DIF. Stepwise purification was then conducted as a diagnostic follow-up, beginning with the temporary exclusion of Item 13, which demonstrated the strongest DIF effect ($\chi^2(6) = 49.63$, $p < .001$). After rerunning the analyses, Items 2, 4 and 5 remained statistically significant. Item 4 ($\chi^2(6) = 49.63$, $p < .001$) and Item 5 ($\chi^2(6) = 19.02$, $p < .001$) were temporarily excluded in subsequent diagnostic steps, with the analyses rerun after each step. Following the temporary exclusion of Item 5, Item 2 remained the only item demonstrating significant DIF in the diagnostic purification model ($\chi^2(6) = 18.78$, p < .01). After these diagnostic purification steps, no remaining items demonstrated significant DIF. These findings suggest that the observed noninvariance at the scale level was primarily driven by a subset of the retained SOD items rather than reflecting pervasive DIF across the entire scale.

Discussion

Study 1 conducted dimensionality checks, IRT modeling and DIF analyses to evaluate the psychometric functioning of the HTW, ASB and SOD measures. Overall, the findings supported the general adequacy of the three measures while also identifying several items that consistently demonstrated weaker psychometric performance across analyses. Specifically, HTW 2 and 3, ASB 5 and 7, and SOD 13 were repeatedly flagged and were therefore removed because of weaker loadings, problematic category functioning, cross-loadings, elevated item or model misfit, or lower item information.

Inspection of item content suggested that the repeatedly flagged items were reverse-worded. Reverse-worded items are often included in self-report scales to reduce acquiescence or response bias; however, accumulating evidence suggests that reverse wording can also introduce measurement problems, particularly when respondents overlook negations, misinterpret item direction, or respond inconsistently across positively and negatively worded items (Barnette, 2000; van Sonderen et al., 2013; Swain et al., 2008; Weijters et al., 2013). These problems may increase random error, reduce item discrimination, and introduce method effects that appear as artificial multidimensionality in factor analytic models. Consistent with this literature, the reverse-worded items in the present analyses tended to show weaker or less stable psychometric functioning than the positively worded items. Therefore, these items were removed to improve the interpretability and measurement quality of the revised HM item pool.

After item reduction, the revised ASB measure contained seven items, the revised HTW measure contained eight items, and the revised SOD measure contained seven items. The ASB items were generally consistent with a unidimensional structure, suggesting that they assessed a coherent adversarial belief system about women and sexual relationships. In contrast, the HTW items showed evidence of two related but distinguishable dimensions. Based on item content, the first HTW dimension appeared to reflect distrust-based hostility toward women, including beliefs that women are deceptive, manipulative, or likely to take advantage of men. The second HTW dimension appeared to reflect resentment-based hostility, including anger, grievance, and antagonism toward women. This distinction suggests that HTW may not be best represented as a single homogeneous construct.

The two-factor structure of HTW was theoretically important for the subsequent studies. Although both distrust- and resentment-based hostility reflect negative attitudes toward women, resentment-based hostility may be more central to the hostile and aggression-relevant component of HM. Whereas distrust-based hostility may reflect suspicion or defensiveness toward women, resentment-based hostility more directly captures anger and antagonism, which are more closely aligned with the Confluence Model's emphasis on hostile, adversarial orientations toward women. For this reason, the HTW resentment dimension was retained as another theoretically meaningful candidate anchor in the bifactor S-1 models tested in Study 2.

The DIF analyses identified statistically significant between-study differences in the functioning of some retained items. However, comparisons of EAP factor scores suggested that, within the present pooled samples, accounting for the detected DIF produced only minimal differences in estimated scores. The affected items were therefore retained rather than removed solely due to statistically significant DIF. This decision, though, should not be interpreted as evidence that the identified DIF is inconsequential across all populations or research settings. The magnitude and practical impact of DIF may differ in other samples, particularly when groups vary more substantially in demographic composition, recruitment context, or response patterns. Future research should therefore continue to examine the invariance and practical consequences of these items before assuming that they function equivalently across populations.

Overall, Study 1 provided an empirically refined item pool for subsequent latent structure analyses. The results supported ASB as a coherent component of HM, clarified the multidimensional structure of HTW, and identified the subset of SOD items most relevant to HM measurement. The results of DIF analyses suggested that between-study noninvariance was localized rather than pervasive, and EAP score comparisons suggested that the practical impact of DIF was minimal. These findings supported the use of the pooled item-level data while also underscoring the need to evaluate item functioning before constructing broader HM measurement models. Study 2 therefore built on these findings by testing how the retained ASB, HTW, and SOD items functioned together within CFA, ESEM, and bifactor S-1 models of hostile masculinity.

CHAPTER THREE

Study 2 (S2)

Method

Participants

The sample for Study 2 used the same pooled dataset as Study 1 ($N = 720$), which combined participants from Project Become 1 and 2. Participant characteristics, recruitment procedures, and eligibility criteria are described in detail in Study 1.

Measures

Hostile masculinity (HM) was assessed using the refined item pool identified in Study 1. This revised item pool retained items that demonstrated stronger overall psychometric functioning and removed items showing weaker performance across dimensionality, item-functioning, item-information, and substantive item-content evaluations. The resulting measure included seven ASB items, eight HTW items, and seven SOD items. All items were rated on a 7-point Likert-type scale ranging from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*), with higher scores indicating greater endorsement of hostile masculinity-related beliefs/attitudes. Consistent with Study 1, the HTW items were represented by two dimensions reflecting distrust-based and resentment-based hostility toward women (i.e., HTW resentment & HTW distrust).

Statistical Analysis

In order to address objective 3, a bifactor-Exploratory Structural Equation Modeling (ESEM) technique was employed to evaluate the psychometric properties of HM, with a particular focus on its dimensionality. HM items, in this framework, are allowed to freely load amongst the specific factors while also loading simultaneously onto the general factors. In Study 2, the hypothesized bifactor S-1 ESEM model was assessed and compared with several alternative specifications including (1) unidimensional, (2) correlated-factor CFA, (3) higher-order CFA, (4) bifactor CFA, (5) bifactor S-1 CFA, (6) correlated-factor ESEM, (7) higher-order ESEM, (8) bifactor ESEM, and (9) bifactor S-1 ESEM models.

Bifactor S-1 Exploratory Structural Equation Modeling. The present study adopted an Exploratory Structural Equation Modeling (ESEM) framework to evaluate the dimensionality of HM using the integrated dataset. ESEM was selected over the traditional factor analytic techniques, such as Confirmatory Factor Analysis (CFA), as it integrates Exploratory Factor Analysis (EFA) into CFA frameworks so that the inherent limitations of either technique can be overcome (Marsh et al., 2014; Schmitt, 2011). More precisely, EFA allows freely estimated cross-loadings and is useful for identifying latent structure. Yet, EFA does not provide a fully specified measurement model within an SEM framework and does not support formal tests of competing theoretical models, measurement invariance, or structural relations among latent variables (Ford et al., 1986; Howard, 2016; Watkins, 2018). CFA, by contrast, imposes the independent clusters model (ICM-CFA), in which items are restricted to load on a single factor with all cross-loadings fixed to zero (McDonald, 1985). For multidimensional constructs with conceptually overlapping components, such restrictions are often unrealistic and can result in inflated factor correlations, distorted loadings, and misleading conclusions regarding dimensionality. Unlike the conventional EFA models, ESEM provides a substantive package of a priori model alternatives that can be incorporated into a testing sequence; and, unlike the conventional CFA models, ESEM allows free cross-loading of items among factors and thus is less restrictive for multidimensional constructs (Marsh et al., 2014). For the current study, the flexibility of ESEM was particularly important for HM and its theoretically distinct yet empirically interrelated subcomponents, SD, HTW, and ASB, allowing a more realistic representation of item-factor relations.

Furthermore, the bifactor S-1 model was selected within the ESEM framework over the hierarchical model and the traditional symmetrical bifactor model. In hierarchical models, each observed indicator is associated with a first-order factor (i.e., SD, HTW, and ASB), and multiple first-order factors are specified as loading on one higher-order factor (i.e., HM). This model structure assumes that the associations between the indicators and the higher-order factor are fully mediated by the first-order factors, resulting in the contribution of the higher-order factor being identical for all indicators within the same first-order factor (Hofmann, 1997; Lewis, 2007; Woltman et al., 2012). In contrast, the symmetrical bifactor model includes a general factor that directly influences all observed indicators, while specific factors capture the unique variance in subsets of indicators, allowing each item to simultaneously load onto both the general and specific factors (Gignac, 2016; Holzinger & Swineford, 1937; Reise, 2012). This approach frees the restrictive proportionality constraint inherent in hierarchical models and is expected to yield better model fit and more accurate construct representation, as supported by previous empirical studies (Alamer, 2022). However, traditional symmetrical bifactor models can introduce interpretational ambiguity when the general factor is defined by all item sets equally, especially when some subdimensions represent more central features of the construct than others. To address these limitations, the bifactor S-1 model designates one theoretically central subdimension as the reference facet that defines the general factor, while the remaining subdimensions are modeled as domain-specific residual factors (Eid et al., 2017, 2018). Conceptual diagrams of the traditional bifactor model and bifactor S-1 model are presented in Figures 10 and 11, respectively.

Figure 10
Conceptual Diagram of a Traditional Bifactor Model

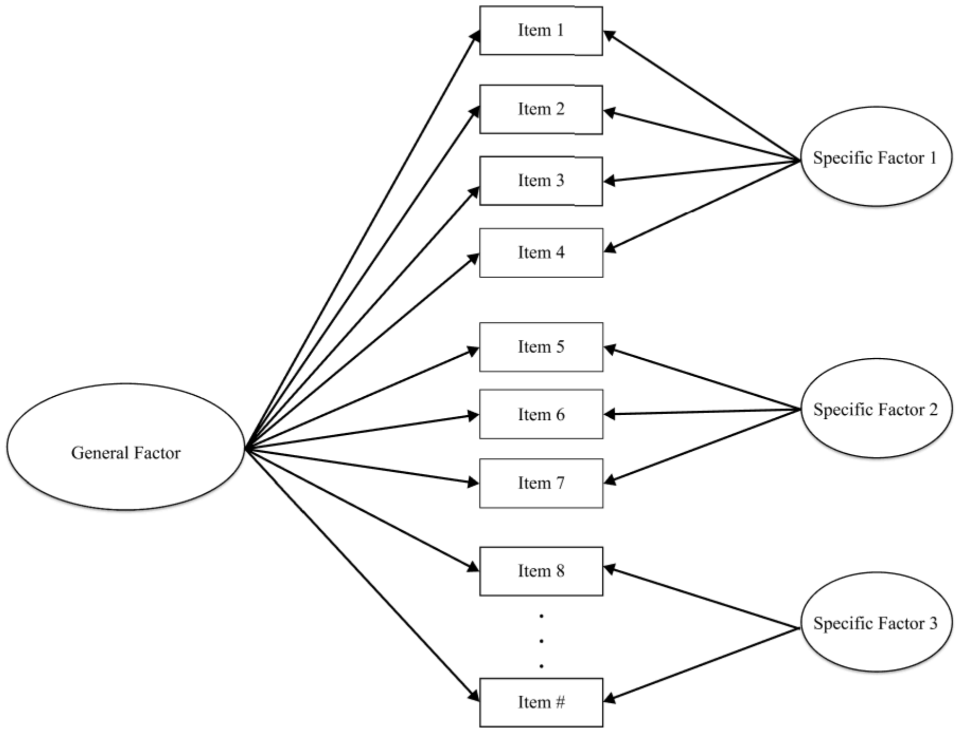
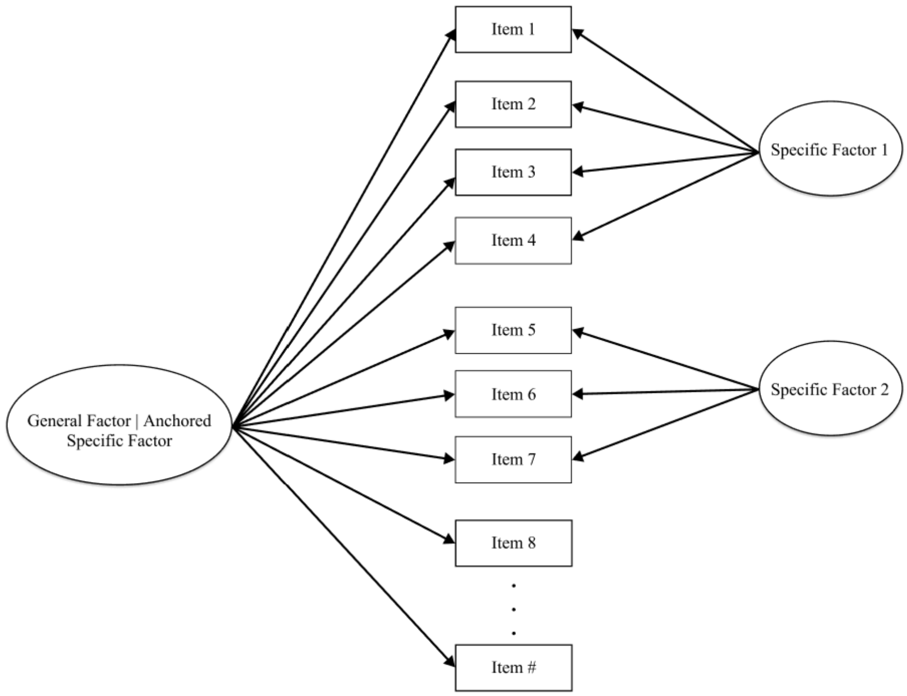


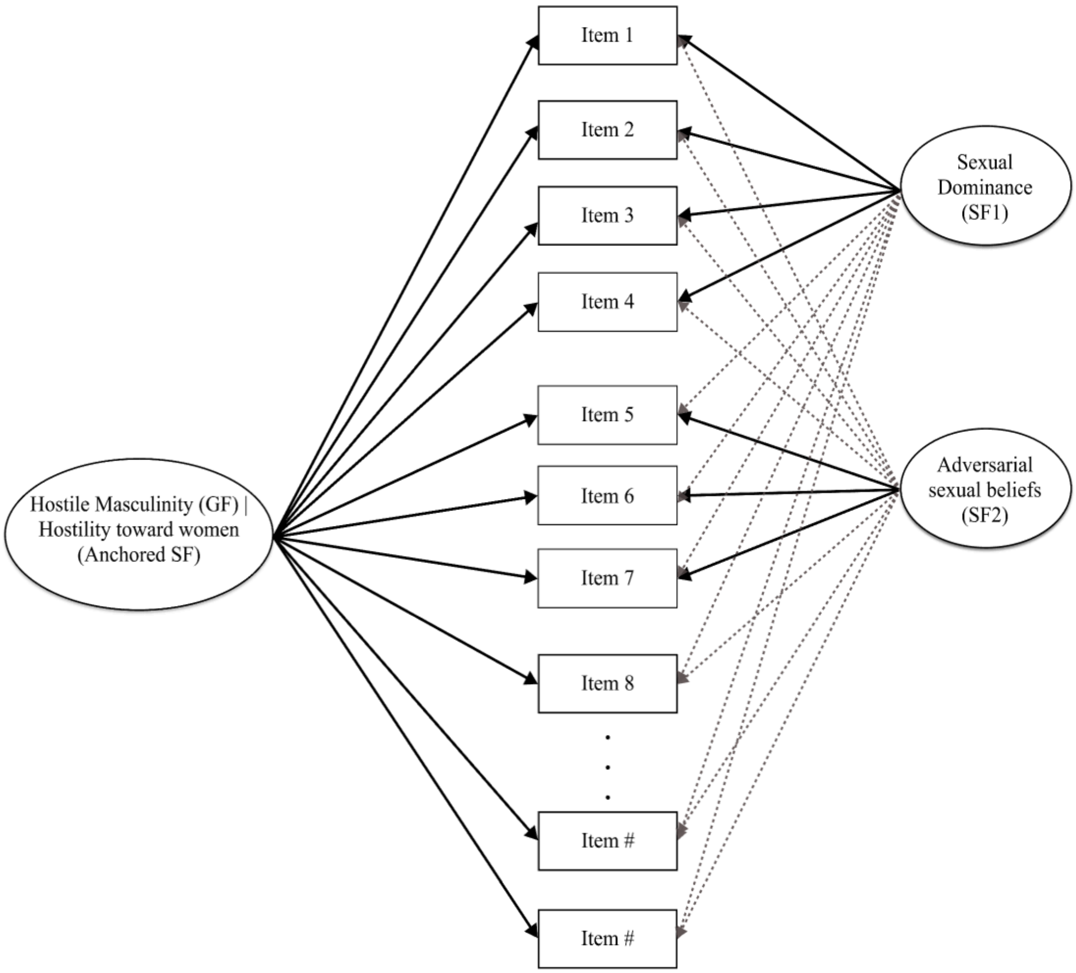
Figure 11
Conceptual Diagram of a Bifactor S-1 Model



The bifactor S-1 model is particularly suited for HM, where items are expected to contribute asymmetrically to the global factor. The "S-1" specification indicates that one of the specific factors serves as the anchor for the global factor, such that items from this reference facet are predicted only by the global factor and not by a specific factor. Because Study 1 supported a two-dimensional structure of HTW and suggested that HM components may contribute differently to the broader construct, three bifactor S-1 anchoring specifications were evaluated. The original HTW-anchored model was retained as an initial theoretically motivated model, given the central role of hostility toward women in the Confluence Model. In addition, HTW resentment and ASB were tested as alternative anchor domains. HTW resentment was examined because Study 1 suggested that resentment-based hostility may more directly capture the affective antagonism central to HM, whereas ASB was examined because adversarial beliefs may represent a broad cognitive orientation linking distrust, resentment, and dominance-related attitudes toward women. In each S-1 specification, items from the anchor domain loaded only on the general HM factor, whereas items from the non-anchor domains loaded on both the general HM factor and their corresponding residual specific factors. This approach allowed the general HM factor to be defined by different theoretically meaningful reference domains and enabled direct comparison of which anchor produced the most stable, parsimonious, and interpretable representation of HM.

Taken together, Study 2 evaluated CFA and ESEM specifications to address Objective 3. ESEM models were included to relax the independent-clusters assumption of CFA by allowing cross-loadings among ASB, HTW, and SOD items. Bifactor S-1 specifications were then used to evaluate whether the general HM factor could be more clearly defined when anchored in a theoretically meaningful reference domain. See Figure 12 for one proposed model for HM. Models were estimated in R using the *lavaan* package (Version 0.6-19; Rosseel, 2012) with robust Maximum Likelihood Estimator (MLR). Model fit was assessed using several fit indices with the following cutoffs: comparative fit index (CFI) and Tucker-Lewis Index (TLI), ≥.95 (Schermelleh-Engel et al., 2003); root mean square error of approximation (RMSEA), ≤ .08 (Browne & Cudeck, 1992); and standardized root mean square residual (SRMR), ≤.08 (Hu & Bentler, 1999). Model comparisons were based on fit indices, information criteria, parameter admissibility, factor-loading patterns, and substantive interpretability, with lower AIC and BIC values indicating stronger relative support for a given model (Raftery, 1995). While traditionally model comparison tests are made using likelihood ratio tests (LRT), there have been multiple issues associated with LRTs, such as sensitivity to sample size (Brannick, 1995) and bias toward the less parsimonious model (for review see Sellbom & Tellegen, 2019). The AIC and BIC indices were both parsimony-adjusted and were the selected indices for model comparison.

Figure 12
Conceptual Diagram of the Bifactor S-1 ESEM Model for HM



In addition to comparing the fit statistics, each factor solution and its overall interpretability were examined and compared using the following criteria: for each first-order model, each observed indicator must show strong primary loadings on one factor ($\lambda \geq .50$) with minimal cross-loadings onto other factors ($\lambda < .30$). Each factor must contain at least five strong loading items. To ensure discriminant validity, the factor intercorrelations should not exceed 0.80. Additional indices including the Explained Common Variance (ECV) and Proportion of Uncontaminated Correlations (PUC) were examined for bifactor models. The ECV indicates the proportion of total common variance explained by the general factor relative to the specific factors, whereas the PUC measures the proportion of correlations that are not influenced by multiple factors. High ECV and PUC provide stronger support for a unidimensional structure, whereas ECV below 0.60 and PUC below .80 is typically the recommended cutoff for establishing unidimensionality (Reise et al., 2013).

Internal Consistency Reliability. The internal consistency of the retained latent factors was estimated by calculating point and interval estimates using McDonald's ω (McDonald, 1999). Traditional estimates of internal consistency are typically calculated with coefficient alpha, whose issues have been previously documented (Dunn et al., 2014). A common problem is the assumption of essentially tau-equivalency required for unbiased estimates (Green et al., 1977). The issue of tau-equivalency is that theoretical "true" score variance does not change across all items. In other words, the responses to items would equally be sensitive to the overall construct. McDonald's ω follows the less restrictive congeneric model and is calculated through the use of the factor loadings of the items, which can be more ideal for the different item sensitivity. Previously, even when tested with a measure that holds essentially-tau equivalency, omega performed similarly to alpha (Zinbarg et al., 2005). After calculating the point estimate for omega, it has been suggested to estimate confidence intervals to gain a greater understanding of the accuracy of the estimate (Raykov, 1998).

For any bifactor or bifactor S-1 model retained, additional reliability indices including omega hierarchical (ωH) and omega hierarchical subscale (ωHS) will also be estimated, as recommended for evaluating reliability in bifactor measurement models (Rodriguez et al., 2016). While McDonald's ω estimates the reliability attributable to all common factors combined, ωH estimates the proportion of reliable variance in the total score attributable to the general factor after accounting for the specific factors, and ωHS estimates the proportion of reliable variance uniquely attributable to each specific factor after controlling for the general factor. These indices were interpreted alongside standardized factor loadings, ECV, PUC, and McDonald's ω to evaluate the extent to which reliable variance is attributable to the general construct versus the specific dimensions. Larger ωH values indicate greater reliable variance retained by the corresponding specific factor beyond the general factor.

Results

Consistent with the analytic plan, a series of CFA and ESEM models were estimated to evaluate the latent structure of hostile masculinity (HM). Given evidence from Study 1 supporting a two-dimensional structure of HTW (i.e., distrust & resentment toward women), the originally proposed plan was expanded to incorporate this revised representation of HTW when applicable. Specifically, the analyses estimated a series of correlated-factor models (i.e., unidimensional, three-factor and four-factor CFA models), two- and three-level higher-order CFA models, bifactor CFA models, and bifactor S-1 CFA models, along with their accompanying four-factor ESEM, higher-order ESEM, bifactor ESEM, and bifactor S-1 ESEM counterparts. Table 10 presents the model fit indices for all estimated models.

Table 10
Comparison of Model Fit Indices for Alternative Measurement Models of Hostile Masculinity

Model	χ^2	*df*	CFI	TLI	RMSEA (90% CI)	SRMR	AIC	BIC	SABIC
Unidimensional Model	1841.96	189	0.772	0.747	.128 [.123, .133]	0.087	51,851.25	52,139.65	51,939.61
Three-factor CFA	869.97	186	0.908	0.896	.082 [.077, .088]	0.045	50,542.78	50,844.92	50,635.35
Four-factor CFA	637.42	183	0.94	0.931	.067 [.061, .073]	0.039	50,243.39	50,559.26	50,340.17
Two-level Higher-order CFA	644.52	185	0.939	0.931	.067 [.061, .073]	0.04	50,247.31	50,554.03	50,341.29
Three-level Higher-order CFA	—	—	—	—	—	—	—	—	—
Bifactor CFA (correlated HTW facets)	477.90	167	0.96	0.95	.057 [.051, .063]	0.032	50,056.68	50,445.80	50,175.90
HTW-anchored S-1 CFA	803.00	176	0.917	0.901	.080 [.075, .086]	0.045	50,463.51	50,811.43	50,570.11
HTW resentment-anchored S-1 CFA	697.17	172	0.932	0.917	.073 [.068, .079]	0.041	50,322.59	50,688.82	50,434.79
ASB-anchored S-1 CFA	603.34	174	0.943	0.931	.067 [.061, .073]	0.033	50,218.71	50,575.79	50,328.12
Four-factor ESEM	427.71	132	0.965	0.944	.065 [.057, .072]	0.022	50,052.74	50,602.08	50,221.05
Two-level Higher-order ESEM	414.75	128	0.964	0.942	.064 [.056, .071]	0.022	50,060.74	50,628.39	50,234.66
Bifactor ESEM	270.90	111	0.982	0.967	.052 [.043, .061]	0.017	49,904.63	50,550.11	50,102.40
HTW-anchored S-1 ESEM	749.21	164	0.923	0.901	.081 [.075, .087]	0.036	50,421.36	50,824.21	50,544.79
HTWF2-anchored S-1 ESEM	451.89	141	0.962	0.943	.062 [.055, .069]	0.024	50,066.65	50,574.80	50,222.34
ASB-anchored S-1 ESEM	573.20	150	0.945	0.924	.072 [.065, .078]	0.031	50,223.48	50,690.42	50,366.54

Note. CFA = confirmatory factor analysis; ESEM = exploratory structural equation modeling; HTW = Hostility Toward Women; ASB = Adversarial Sexual Beliefs; RMSEA values are reported with 90% confidence intervals. For the three-level higher-order CFA, χ^2, CFI, TLI, and RMSEA were unavailable

Confirmatory Factor Analysis Models

Within the CFA framework, three models, including a unidimensional model, a three-factor model, and a four-factor model, were first estimated to evaluate the multidimensionality of the HM. The unidimensional model demonstrated poor model fit to the data, $\chi^2(189) = 1841.96$, CFI = .772, TLI = .747, RMSEA = .128, SRMR = .087. In contrast, the three-factor CFA model specifying HTW, ASB, and SOD as correlated latent variables demonstrated substantially improved fit, $\chi^2(186) = 869.97$, CFI = .908, TLI = .896, RMSEA = .082, SRMR = .045. The four-factor CFA model, which further separated HTW into two distrust and resentment factors, provided the best fit among the three, $\chi^2(183) = 637.42$, CFI = .940, TLI = .931, RMSEA = .067, SRMR = .039. Standardized factor loadings for each model are presented in Table 11. These findings supported the multidimensional structure of HM and indicated that HTW was better represented by two correlated dimensions. Moreover, latent factor correlations in the four-factor CFA model ranged from .512 to .872 (see Table 12), with the strongest correlation observed between ASB and HTW distrust ($r = .872$) and between ASB and HTW resentment ($r = .828$). Given these strong intercorrelations, higher-order and bifactor models were subsequently estimated to evaluate whether the associations among the ASB, HTW and SOD could be accounted for by a broader HM construct.

Table 11

Standardized Factor Loadings for the One-Factor, Three-Factor, and Four-Factor HM Models

One-factor Model		Three-factor Model			Four-factor Model		
Item	Loading	Factor	Item	Loading	Factor	Item	Loading
ASB1	0.754	HTW =~	HTW5	0.812	HTW_F1 =~	HTW5	0.852
ASB2	0.689		HTW6	0.8		HTW6	0.839
ASB3	0.769		HTW4	0.748		HTW4	0.799
ASB4	0.734		HTW7	0.744		HTW1	0.673
ASB6	0.806		HTW8	0.727	HTW_F2 =~	HTW9	0.802
ASB8	0.807		HTW9	0.727		HTW10	0.806
ASB9	0.758		HTW10	0.727		HTW7	0.797
HTW1	0.652		HTW1	0.657		HTW8	0.77
HTW4	0.724	SOD =~	SOD5	0.827	ASB =~	ASB6	0.823
HTW5	0.78		SOD15	0.785		ASB8	0.822
HTW6	0.758		SOD7	0.782		ASB3	0.797
HTW7	0.707		SOD11	0.71		ASB1	0.774
HTW8	0.701		SOD4	0.68		ASB9	0.773
HTW9	0.694		SOD2	0.605		ASB4	0.753
HTW10	0.703	ASB =~	ASB6	0.824		ASB2	0.706
SOD2	0.404		ASB8	0.821	SOD =~	SOD5	0.827
SOD4	0.515		ASB3	0.797		SOD15	0.785
SOD5	0.564		ASB1	0.774		SOD7	0.783
SOD7	0.464		ASB9	0.773		SOD11	0.71
SOD11	0.521		ASB4	0.752		SOD4	0.68
SOD15	0.512		ASB2	0.707		SOD2	0.605

Note. Values are standardized factor loadings. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance. In the one-factor model, all items loaded onto a single general HM factor. In the three-factor model, items loaded onto HTW, ASB, and SOD factors. In the four-factor model, HTW items were separated into two factors. Blank cells indicate that the factor label continues from the row above.

Table 12

Latent Factor Correlations for the Four-Factor CFA Model

Factor	1	2	3	4
1. ASB	—			
2. SOD	0.611	—		
3. HTW distrust	0.872	0.513	—	
4. HTW resentment	0.828	0.512	0.789	—

Note. ASB = Adversarial Sexual Beliefs; SOD = Sexual Dominance; HTW = Hostility Toward Women

A three-level higher-order CFA model was specified, in which HTW distrust and resentment loaded onto a second-order HTW factor, and HTW, ASB and SOD loaded onto a third-order HM factor. However, the model failed to converge and could not be interpreted. The more parsimonious two-level higher-order model was then retained, with HTW distrust, HTW resentment, ASB and SOD loading directly onto the second-order HM. This two-level model demonstrated acceptable fit, $\chi^2(185) = 644.52$, CFI = .939, TLI = .931, RMSEA = .067, SRMR = .040. Table 13 presents the standardized factor loadings of the two-level higher-order CFA model. At the higher-order level, ASB demonstrated the strongest loading on HM (λ = .971), followed by HTW distrust (λ = .899), HTW resentment (λ = .858), and SOD (λ = .610). At the first-order level, all indicators loaded substantially on their intended factors, with standardized loadings ranging from .604 to .851, which also provided support for the reliability of the retained factor structure. Despite its acceptable fit, the two-level higher-order model was not retained as the final model because hierarchical models do not directly partition variance into general and domain-specific components. Bifactor models offer greater flexibility by allowing indicators to load directly on a general factor while also estimating residual-specific factors, making them better suited for evaluating the multidimensional structure of HM. Therefore, bifactor models were subsequently estimated.

Table 13
Standardized Loadings for the Two-Level Higher-Order CFA Model

Panel A. Higher-Order Factor Loadings

Higher-Order Factor	Lower-Order Factor	Loading
HM =~	ASB	0.971
	HTW_F1	0.899
	HTW_F2	0.858
	SOD	0.61

Panel B. First-Order Factor Loadings

Factor	Item	Loading
ASB =~	ASB6	0.824
	ASB8	0.823
	ASB3	0.796
	ASB9	0.774
	ASB1	0.772
	ASB4	0.753
	ASB2	0.707
SOD =~	SOD5	0.828
	SOD15	0.784
	SOD7	0.782
	SOD11	0.71
	SOD4	0.681
	SOD2	0.604
HTW distrust =~	HTW5	0.851
	HTW6	0.837
	HTW4	0.8
	HTW1	0.674
HTW resentment =~	HTW10	0.807
	HTW9	0.802
	HTW7	0.796
	HTW8	0.77

Note. Values are standardized factor loadings from the two-level higher-order CFA model. HM = hostile masculinity; ASB = adversarial sexual beliefs; HTW = hostility toward women; SOD = sexual dominance. Panel A presents loadings of the lower-order factors on the higher-order HM factor. Panel B presents item loadings on their corresponding lower-order factors.

Moreover, a bifactor CFA model was estimated in which all items loaded simultaneously onto a general HM factor and four orthogonal specific factors representing ASB, SOD, HTW distrust and HTW resentment. Consistent with the conceptualization of HTW as two dimensions of the same scale, the HTW distrust and resentment specific factors were allowed to correlate (r = .243). As hypothesized, the bifactor CFA model demonstrated significantly improved fit relative to the two-level higher-order CFA model, $\Delta\chi^2(18) = 149.10$, $p < .001$, $\chi^2(167) = 477.90$, CFI = .960, TLI = .950, RMSEA = .057, SRMR = .032. Inspection of standardized loadings indicated that ASB items loaded strongly on the general HM factor (λs = .692–.854), whereas residual loadings on the ASB-specific factor were relatively weak (λs = -.139–.413; see Table 14). This finding suggested that much of the variance in ASB items was accounted for by the general HM dimension. Comparatively, the HTW distrust, HTW resentment, and especially SOD specific factors retained more unique variance after the shared HM variance was extracted. These findings together supported the presence of a meaningful general HM factor while suggesting that among the four specific factors, ASB was most strongly associated with the general construct. SOD, HTW distrust, and HTW resentment retained substantial domain-specific variance beyond that explained by the general HM factor.

Table 14

Standardized Loadings for the Bifactor CFA (correlated HTW distrust & HTW resentment)

General Factor	Item	Factor Loading	Specific Factors	Item	Factor Loading
HM =~	ASB8	**0.854**	HTW_s1 =~	HTW5	**0.431**
	ASB6	**0.807**		HTW6	**0.429**
	ASB9	**0.805**		HTW4	**0.383**
	ASB3	**0.773**		HTW1	0.186
	ASB1	**0.748**	HTW_s2 =~	HTW9	**0.515**
	ASB4	**0.748**		HTW10	**0.462**
	ASB2	**0.692**		HTW7	**0.449**
	HTW5	**0.742**		HTW8	**0.377**
	HTW6	**0.728**	ASB =~	ASB3	**0.413**
	HTW4	**0.699**		ASB1	**0.397**
	HTW10	**0.66**		ASB2	0.132
	HTW7	**0.657**		ASB6	0.111
	HTW8	**0.657**		ASB4	0.053
	HTW1	**0.644**		ASB8	−.139
	HTW9	**0.637**		ASB9	−.165
	SOD5	**0.5**	SOD =~	SOD7	**0.702**
	SOD11	**0.473**		SOD15	**0.654**
	SOD4	**0.465**		SOD5	**0.644**
	SOD15	**0.457**		SOD11	**0.525**
	SOD7	**0.403**		SOD2	**0.488**
	SOD2	**0.354**		SOD4	**0.477**

Note. HM = hostile masculinity; HTW = hostility toward women; HTW_s1 = HTW distrust residual factor; HTW_s2 = HTW resentment residual factor; ASB_s = Adversarial Sexual Beliefs residual factor; SOD_s = Sexual Dominance residual factor. The general factor (HM) was specified as orthogonal to all specific factors, and all specific factors were constrained to be orthogonal except for HTW_s1 and HTW_s2, which were allowed to correlate. The estimated correlation between HTW_s1 and HTW_s2 was r = .243.

Bifactor S-1 CFA Models

Because ASB items exhibited particularly strong loadings on the general HM factor and HTW showed a two-factor structure, ASB and HTW resentment were estimated as alternative anchor domains in addition to the original proposed HTW-anchored S-1 model. As a result, the HTW-anchored S-1 CFA model provided acceptable fit, $\chi^2(176) = 803.00$, $p < .001$, CFI = .917, TLI = .901, RMSEA = .080, SRMR = .045, but did not demonstrate improved fit relative to the four-factor, higher-order or bifactor CFA models. Inspection of the standardized factor loadings (See Table 15) indicated that the general factor was defined not only by HTW items (λs = .672–.806), but also several ASB items, which loaded more strongly on the general HM (λs = .653–.803) than the residual ASB-specific factor (λs = .082–.579). In contrast, SOD items showed comparatively weaker general-factor loadings (λs = .341–.482) and consistently retained stronger loadings on their residual specific factor(λs = .491–.716).

Table 15

Standardized Loadings for the HTW-anchored S-1 CFA model

Factor	Item	Factor Loading	Factor	Item	Factor Loading
HM_HTW =~	ASB1	0.683	ASB_s =~	ASB3	0.579
	ASB2	0.653		ASB1	0.465
	ASB3	0.687		ASB2	0.251
	ASB4	0.706		ASB6	0.239
	ASB6	0.779		ASB4	0.221
	ASB8	0.803		ASB8	0.116
	ASB9	0.76		ASB9	0.082
	HTW1	0.672	SOD_s =~	SOD7	0.716
	HTW4	0.752		SOD15	0.675
	HTW5	0.806		SOD5	0.657
	HTW6	0.799		SOD11	0.537
	HTW7	0.733		SOD2	0.497
	HTW8	0.715		SOD4	0.491
	HTW9	0.707			
	HTW10	0.712			
	SOD2	0.341			
	SOD4	0.451			
	SOD5	0.482			
	SOD7	0.378			
	SOD11	0.46			
	SOD15	0.427			

Note. HM_HTW = general hostile masculinity factor anchored by Hostility Toward Women (HTW); ASB_s = residual Adversarial Sexual Beliefs factor; SOD_s = residual Sexual Dominance factor. The general factor was specified as orthogonal to the specific factors, and the specific factors were constrained to be orthogonal to one another.

The HTW resentment-anchored S-1 model showed considerably improved fit, $\chi^2(172) = 697.17$, $p < .001$, CFI = .932, TLI = .917, RMSEA = .073, SRMR = .041. In this model, ASB items continued to load strongly on the general factor, with three items among the highest-loading indicators (See Table 16). The residual HTW distrust factor was more modestly defined (λs = .180 to .402), whereas the residual SOD factor remained clearly identifiable (λs = .487 to .712).

Table 16
Standardized Loadings for the HTW resentment-anchored S-1 CFA model

Factor	Item	Factor Loading	Factor	Item	Factor Loading
HM_HTW resentment	ASB1	0.696	HTW distrust_s	HTW4	0.402
	ASB2	0.672		HTW5	0.385
	ASB3	0.705		HTW6	0.375
	ASB4	0.702		HTW1	0.18
	ASB6	0.792	ASB_s	ASB3	0.564
	ASB8	0.801		ASB1	0.444
	ASB9	0.765		ASB4	0.213
	HTW1	0.647		ASB2	0.209
	HTW4	0.699		ASB6	0.204
	HTW5	0.761		ASB8	0.096
	HTW6	0.753		ASB9	0.05
	HTW7	0.748	SOD_s	SOD7	0.712
	HTW8	0.73		SOD15	0.668
	HTW9	0.729		SOD5	0.652
	HTW10	0.738		SOD11	0.533
	SOD2	0.35		SOD2	0.491
	SOD4	0.455		SOD4	0.487
	SOD5	0.49			
	SOD7	0.387			
	SOD11	0.465			
	SOD15	0.437			

Note. HM_HTW = general hostile masculinity factor anchored by Hostility Toward Women (HTW); ASB_s = residual Adversarial Sexual Beliefs factor; SOD_s = residual Sexual Dominance factor. The general factor was specified as orthogonal to the specific factors, and the specific factors were constrained to be orthogonal to one another.

The ASB-anchored S-1 model provided the best fit among these three S-1 specifications, $\chi^2(174) = 603.34$, $p < .001$, CFI = .943, TLI = .931, RMSEA = .067, SRMR = .033. Because HTW distrust and HTW resentment were derived from the same HTW scale, their residual specific factors were allowed to correlate in this model, $r = .237$. As indicated by the standardized factor loadings, the general factor was especially strongly defined by ASB items (λs = .707–.824), and HTW items also showed moderate to strong general-factor loadings (see Table 17). The residual HTW distrust and resentment factors retained smaller but meaningful loadings, whereas SOD continued to load more strongly on its residual specific factor (λs = .467–.697) than on the general factor (λs = .367–.512).

Table 17
Standardized Loadings for the ASB-anchored S-1 CFA model

Factor	Item	Factor Loading	Factor	Item	Factor Loading
HM_ASB	ASB1	0.773	HTW_s1	HTW6	0.44
	ASB2	0.707		HTW5	0.427
	ASB3	0.795		HTW4	0.406
	ASB4	0.752		HTW1	0.22
	ASB6	0.82	HTW_s2	HTW9	0.5
	ASB8	0.824		HTW10	0.451
	ASB9	0.775		HTW7	0.448
	HTW1	0.629		HTW8	0.369
	HTW4	0.688	SOD_s	SOD7	0.697
	HTW5	0.742		SOD15	0.647
	HTW6	0.722		SOD5	0.634
	HTW7	0.658		SOD11	0.52
	HTW8	0.662		SOD2	0.481
	HTW9	0.647		SOD4	0.467
	HTW10	0.667			
	SOD2	0.364			
	SOD4	0.475			
	SOD5	0.512			
	SOD7	0.412			
	SOD11	0.478			
	SOD15	0.468			

Note. HM_ASB = general hostile masculinity factor anchored by Adversarial Sexual Beliefs (ASB); HTW_s1 = residual HTW distrust factor; HTW_s2 = residual HTW resentment factor; SOD_s = residual Sexual Dominance factor. The general factor was specified as orthogonal to all specific factors. The residual HTW factors (HTW_s1 and HTW_s2) were permitted to correlate (r = .237), whereas all other specific factors were constrained to be orthogonal.

Taken together, these results indicated that ASB functioned as the strongest anchor for the general HM factor, whereas SOD represented the most distinct residual domain, supporting the ASB-anchored S-1 model as the most empirically coherent representation of the data, and HM was primarily reflected in a broad ASB-centered general factor with residual components for SOD, HTW distrust and HTW resentment.

Exploratory Structural Equation Modeling

To free the restrictive zero cross-loading assumptions imposed by CFA, ESEM models were estimated, including a four-factor ESEM model, a higher-order ESEM model, a bifactor ESEM model, and three bifactor S-1 ESEM models. Compared to its corresponding CFA model, the four-factor ESEM demonstrated improved fit, $\chi^2(132) = 427.71$, $p < .001$, CFI = .965, TLI = .944, RMSEA = .065, SRMR = .022, with most items loading most strongly on their intended factors. However, multiple ASB items showed notable cross-loadings on the HTW distrust factor, and the latent factor correlations were also moderate to strong (See Table 18), suggesting the need to evaluate whether the shared variance among these four factors could be represented by a more integrative model. Table 19 presents latent factor correlations of the four-factor ESEM model.

Table 18

Standardized Loadings for the ESEM Four-Factor Model

Item	F1 (ASB)	F3 (HTW distrust)	F4 (HTW resentment)	F2 (SOD)
ASB1	**0.766**	0.012	0.036	0.055
ASB2	**0.48**	0.145	0.158	-0.005
ASB3	**0.933**	-0.07	-0.001	0.019
ASB4	**0.433**	**0.439**	-0.083	0.02
ASB6	**0.48**	0.26	0.18	-0.006
ASB8	**0.313**	**0.607**	-0.023	-0.012
ASB9	0.266	**0.542**	0.051	-0.026
HTW1	-0.026	**0.637**	0.054	0.085
HTW4	0.013	**0.79**	-0.026	0.04
HTW5	0.093	**0.654**	0.128	0.015
HTW6	-0.001	**0.77**	0.123	-0.045
HTW7	-0.005	0.203	**0.642**	0.05
HTW8	0.161	0.155	**0.512**	0.026
HTW9	0.167	-0.009	**0.72**	-0.003
HTW10	0.256	0.003	**0.635**	-0.03
SOD2	-0.011	0.019	0.027	**0.59**
SOD4	0.112	0.065	-0.003	**0.567**
SOD5	0.086	-0.01	0.029	**0.761**
SOD7	-0.008	-0.026	-0.031	**0.833**
SOD11	-0.01	0.126	0.017	**0.645**
SOD15	0.048	0.005	-0.028	**0.777**

Note. F1 corresponds primarily to Adversarial Sexual Beliefs (ASB), F2 to Sexual Dominance (SOD), F3 to the HTW distrust factor, and F4 to the HTW resentment factor.

Table 19

Latent Factor Correlations for the Four-Factor ESEM Model

Factor	1	2	3	4
1. ASB	—			
2. SOD	0.553	—		
3. HTW distrust	0.7	0.444	—	
4. HTW resentment	0.605	0.39	0.635	—

Note. ASB = Adversarial Sexual Beliefs; SOD = Sexual Dominance; HTW = Hostility Toward Women. Values are standardized latent factor correlations from the four-factor ESEM model.

The two-level higher-order ESEM model was then tested and demonstrated adequate global fit, $\chi^2(128) = 414.75$, $p < .001$, CFI = .964, TLI = .942, RMSEA = .064, 90% CI [.056, .071], SRMR = .022. Despite the robust fit, the model yielded an inadmissible solution with negative latent variance estimates. The factor loadings were therefore not interpreted. Similarly, the bifactor ESEM model showed superior overall fit, $\chi^2(111) = 270.90$, $p < .001$, CFI = .982, TLI = .967, RMSEA = .052, SRMR = .017, and yielded interpretable standardized loadings. Still, it produced a non-positive definite parameter covariance matrix, suggesting instability in the parameter estimates. Inspection of the standardized factor loadings (see Table 20) indicated that the general factor was only modestly to moderately defined, with only two items loading above .50. Among the specific factors, S1 showed strong loadings but was broadly defined by both ASB and HTW items rather than representing a clearly distinct domain-specific factor. S3 and S4 were primarily defined by SOD and HTW resentment items, respectively, whereas S2 lacked a clear substantive interpretation. The pattern of this model, similar to the bifactor CFA model, suggested that a conventional bifactor specification may have produced an unstable or difficult-to-interpret partitioning of general and domain-specific variance. The results also pointed to the use of S-1 models to evaluate whether the general HM factor could be more clearly defined when anchored in a theoretically and empirically meaningful reference domain.

Table 20
Standardized Loadings for the Bifactor ESEM

Item	HM	S1	S2	S3	S4
ASB1	**0.419**	**0.731**	**0.362**	0.05	0.005
ASB2	**0.319**	**0.676**	0.187	-0.016	0.085
ASB3	**0.458**	**0.748**	**0.466**	0.016	-0.016
ASB4	**0.352**	**0.757**	0.11	-0.006	-0.112
ASB6	0.297	**0.822**	0.164	-0.008	0.096
ASB8	**0.356**	**0.844**	0.004	-0.033	-0.084
ASB9	**0.301**	**0.793**	-0.005	-0.042	-0.021
SOD2	0.21	**0.342**	-0.012	**0.47**	0.016
SOD4	**0.604**	**0.38**	-0.012	**0.467**	-0.046
SOD5	**0.554**	**0.427**	-0.001	**0.65**	-0.001
SOD7	0.171	**0.408**	0.014	**0.682**	-0.008
SOD11	0.21	**0.48**	-0.034	**0.525**	0.003
SOD15	-0.077	**0.525**	0.049	**0.743**	0.009
HTW1	**0.305**	**0.674**	-0.183	0.042	-0.03
HTW4	**0.403**	**0.763**	-0.21	-0.004	-0.112
HTW5	**0.366**	**0.811**	-0.141	-0.009	0.016
HTW6	**0.312**	**0.814**	-0.211	-0.054	0.006
HTW7	0.299	**0.684**	-0.121	0.031	**0.421**
HTW8	**0.37**	**0.662**	-0.021	0.008	**0.326**
HTW9	**0.358**	**0.645**	0.004	-0.014	**0.488**
HTW10	**0.318**	**0.671**	0.06	-0.034	**0.43**

Note. HM = hostile masculinity; HTW = hostility toward women; HTW_s1 = HTW distrust residual factor; HTW_s2 = HTW resentment residual factor; ASB_s = Adversarial Sexual Beliefs residual factor; SOD_s = Sexual Dominance residual factor. The general factor (HM) was specified as orthogonal to all specific factors, and all specific factors were allowed to correlate. Factor loadings ≥ .30 were bolded

Bifactor S-1 ESEM

Three bifactor S-1 ESEM models were estimated using HTW, HTW resentment, and ASB as anchor domains. The HTW-anchored model demonstrated acceptable but comparatively weaker fit, $\chi^2(164) = 749.21$, $p < .001$, CFI = .923, TLI = .901, RMSEA = .081, SRMR = .036. Largely consistent with its corresponding CFA model, HTW and ASB items loaded strongly on the general HM factor, whereas SOD items showed weaker general-factor loadings and stronger loadings on the residual SOD-specific factor. The residual ASB-specific factor was also weakly clearly defined. With residual specific-factor correlations freely estimated in ESEM, the residual ASB- and SOD-specific factor remained moderately correlated, $r = .343$, suggesting that shared variance remained among the residual domains. Table 21 presented factor loadings for the HTW-anchored S-1 ESEM model.

Table 21

Standardized Loadings for the HTW-Anchored Bifactor S-1 ESEM

Item	HM_HTW	ASB_s	SOD_s
HTW1	0.673		
HTW4	0.753		
HTW5	0.807		
HTW6	0.804		
HTW7	0.735		
HTW8	0.714		
HTW9	0.707		
HTW10	0.712		
ASB1	0.671	0.476	0.032
ASB2	0.648	0.274	-0.022
ASB3	0.674	0.59	0
ASB4	0.7	0.234	0.006
ASB6	0.774	0.261	-0.018
ASB8	0.801	0.145	-0.025
ASB9	0.759	0.115	-0.039
SOD2	0.332	-0.025	0.512
SOD4	0.433	0.06	0.485
SOD5	0.463	0.046	0.656
SOD7	0.364	-0.026	0.731
SOD11	0.45	-0.027	0.556
SOD15	0.412	0.012	0.679

Note. Values are standardized factor loadings from the HTW-anchored bifactor S-1 exploratory structural equation modeling (ESEM) model. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance. In this model, HTW served as the reference facet defining the general HM factor; therefore, HTW items load only on the general HM factor. ASB Specific and SOD Specific represent residual specific factors after accounting for the general HM factor. Dashes indicate parameters that were fixed to zero by model specification. The general HM factor was orthogonal to both specific factors. The ASB Specific and SOD Specific factors were positively correlated, $r = .343$.

The HTW resentment-anchored model demonstrated the best global fit among the three S-1 ESEM models, $\chi^2(141) = 451.89$, $p < .001$, CFI = .962, TLI = .943, RMSEA = .062, SRMR = .024. The general HM factor was strongly defined by the HTW resentment anchor items and also showed moderate to strong loadings from ASB and HTW distrust items. SOD items again loaded more strongly on the residual SOD-specific factor than on the general factor (see Table 22. However, the residual ASB-specific factor was defined by only a small subset of ASB items, and residual correlations indicated remaining overlap among the specific factors, particularly between ASB and SOD (r = .376) and between ASB and HTW distrust (r = .305). See Table 23 for the residual factor correlations among the specific factors.

Table 22

Standardized Loadings for the HTW Resentment-Anchored Bifactor S-1 ESEM

Item	HM_HTW resentment	ASB_s	SOD_s	HTW distrust_s
HTW7	0.798			
HTW8	0.768			
HTW9	0.804			
HTW10	0.8			
HTW1	0.545	-0.029	0.088	0.427
HTW4	0.591	0.008	0.051	0.535
HTW5	0.688	0.045	0.022	0.425
HTW6	0.676	-0.015	-0.031	0.489
ASB1	0.642	0.506	0.039	0.007
ASB2	0.617	0.309	-0.015	0.08
ASB3	0.651	0.616	0.008	-0.036
ASB4	0.579	0.299	0.022	0.301
ASB6	0.725	0.297	-0.01	0.157
ASB8	0.66	0.215	-0.005	0.4
ASB9	0.639	0.173	-0.018	0.359
SOD2	0.312	-0.018	0.523	0.002
SOD4	0.4	0.063	0.503	0.035
SOD5	0.442	0.041	0.674	-0.013
SOD7	0.335	-0.019	0.744	-0.013
SOD11	0.407	-0.015	0.572	0.074
SOD15	0.379	0.027	0.689	-0.003

Note. Values are standardized factor loadings from the HTW resentment-anchored bifactor S-1 exploratory structural equation modeling (ESEM) model. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance. HTW resentment served as the reference facet defining the general HM factor; therefore, HTW resentment items load only on the general HM factor. ASB Specific, SOD Specific, and HTW distrust Specific represent residual specific factors after accounting for the general HM factor.

Table 23

Latent Factor Correlations for the HTW Resentment-Anchored Bifactor S-1 ESEM Model

Factor	HM	ASB_s	SOD_s	HTW_F1_s
HM	1			
ASB_s	0	1		
SOD_s	0	0.376	1	
HTW distrust_s	0	0.305	0.144	1

Note. Values are latent factor correlations from the HTW distrust-anchored bifactor S-1 ESEM model. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance. The general HM factor was specified as orthogonal to the residual specific factors. The specific factors were freely correlated.

The ASB-anchored model demonstrated acceptable fit, $\chi^2(150) = 573.20$, $p < .001$, CFI = .945, TLI = .924, RMSEA = .072, SRMR = .031. Although this model did not provide the best global fit, its loading and residual-correlation patterns were more substantively coherent. The general HM factor was strongly defined by ASB items and also showed moderate to strong loadings from HTW items. SOD items continued to show stronger loadings on the residual SOD-specific factor than on the general factor (see Table 24). In addition, residual factor correlations were generally small, with only the residual HTW distrust and HTW resentment factors showing a meaningful association, $r = .237$; the remaining residual correlations were near zero (see Table 26). This pattern suggested that once ASB-based general variance was extracted, relatively little shared variance remained among the residual domains. Therefore, although the HTW resentment-anchored model showed the strongest global fit, the ASB-anchored S-1 ESEM model provided the clearest partitioning of general HM variance and residual domain-specific variance.

Table 24

Standardized Loadings for the ASB-Anchored Bifactor S-1 ESEM

Item	HM_ASB	SOD_s	HTW distrust_s	HTW resentment_s
ASB1	0.772			
ASB2	0.705			
ASB3	0.795			
ASB4	0.753			
ASB6	0.823			
ASB8	0.823			
ASB9	0.775			
HTW1	0.627	0.027	0.237	-0.02
HTW4	0.688	0.003	0.439	-0.078
HTW5	0.741	-0.012	0.407	0.014
HTW6	0.723	-0.075	0.432	0.002
HTW7	0.651	0.019	0.099	0.426
HTW8	0.658	0.012	0.058	0.358
HTW9	0.649	-0.005	-0.002	0.5
HTW10	0.675	-0.03	-0.065	0.469
SOD2	0.364	0.482	-0.009	0.013
SOD4	0.463	0.482	0.077	0.008
SOD5	0.505	0.645	0.048	0.027
SOD7	0.422	0.687	-0.029	-0.02
SOD11	0.474	0.524	0.06	0.003
SOD15	0.478	0.634	-0.036	-0.037

Note. Values are standardized factor loadings from the ASB-anchored bifactor S-1 exploratory structural equation modeling (ESEM) model. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance. HM_ASB = the ASB-anchored HM factor: ASB served as the reference facet defining the general HM factor; therefore, ASB items load only on the general HM factor. SOD_s = the specific factor of SOD; HTW distrust_s = the specific factor of HTW distrust; HTW resentment_s = the specific factor of HTW resentment.

Table 25

Latent Factor Correlations for the ASB-Anchored Bifactor S-1 ESEM Model

Factor	HM	ASB_s	SOD_s	HTW_F1_s
HM	1			
SOD_s	0	1		
HTW distrust_s	0	-0.018	1	
HTW resentment_s	0	0.014	0.237	1

Note. Values are latent factor correlations from the ASB-anchored bifactor S-1 ESEM model. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance. The general HM factor was specified as orthogonal to the residual specific factors. The specific factors were freely correlated.

Explained Common Variance and Proportion of Uncontaminated Correlations

Explained common variance (ECV) and percentage of uncontaminated correlations (PUC) were examined to evaluate the relative strength of the general factor across the bifactor and bifactor S-1 models (see Table 26). Among the CFA models, ECV values ranged from .702 to .758, suggesting that the general HM factor accounted for a substantial proportion of common variance across different specifications. The HTW-anchored S-1 CFA model produced the highest ECV (ECV = .758), followed by the ASB-anchored S-1 CFA model (ECV = .729), indicating relatively strong general-factor saturation in both models. In contrast, the bifactor ESEM model yielded a markedly lower ECV (ECV = .157), despite demonstrating the strongest global fit. This discrepancy, consistent with the model's factor-loading pattern, suggested that the improved global fit was driven less by a dominant general HM factor than by the redistribution of common variance across exploratory specific factors.

Table 26
ECV and PUC for Bifactor and Bifactor S-1 Models of Hostile Masculinity

Model	General Factor	ECV	PUC
Bifactor CFA (correlated HTW facets)	HM	0.702	0.771
HTW-anchored bifactor S-1 CFA	HM_HTW	0.758	0.829
HTW resentment-anchored bifactor S-1 CFA	HM_HTWF2	0.736	0.8
ASB-anchored bifactor S-1 CFA	HM_ASB	0.729	0.871
Bifactor ESEM (four specific factors)	HM	0.157	0.771
HTW-anchored bifactor S-1 ESEM	HM_HTW	0.746	0.829
HTW resentment-anchored bifactor S-1 ESEM	HM_HTWF2	0.629	0.8
ASB-anchored bifactor S-1 ESEM	HM_ASB	0.726	0.871

Note. ECV = Explained Common Variance; PUC = Proportion of Uncontaminated Correlations; HM = Hostile Masculinity; HTW = Hostility Toward Women; ASB = Adversarial Sexual Beliefs; CFA = confirmatory factor analysis; ESEM = exploratory structural equation modeling.

Among the S-1 ESEM models, ECV values ranged from .629 to .746. The HTW-anchored S-1 ESEM model produced the highest ECV (ECV = .746), followed by the ASB-anchored S-1 ESEM model (ECV = .726), whereas the HTW resentment-anchored S-1 ESEM model yielded a lower ECV (ECV = .629). PUC values ranged from .771 to .871 across models, suggesting that a large proportion of item correlations reflected general-factor variance uncontaminated by specific-factor membership. Taken together, these findings suggest that although ESEM specifications generally improved global fit, the strength and interpretability of the general HM factor depended substantially on model parameterization and anchor selection.

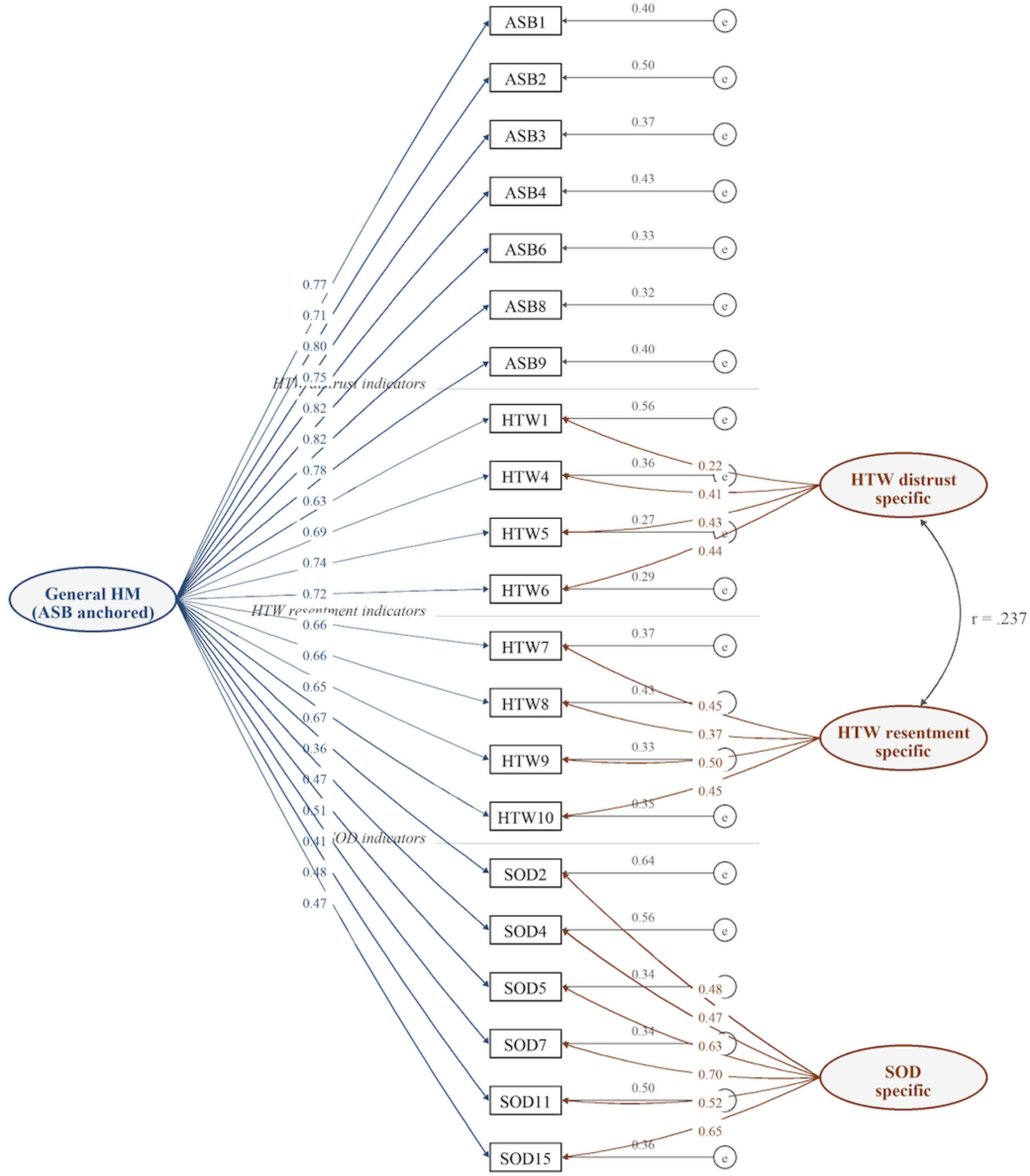
ASB-anchored S-1 CFA VS ESEM

Given the stability and interpretability of the ASB-anchored general factor across CFA and ESEM frameworks, the ASB-anchored S-1 CFA and ESEM models were directly compared to guide final model selection. Although the ASB-anchored S-1 ESEM model showed slightly higher CFI and lower SRMR than the corresponding CFA model, the differences were minimal (ΔCFI = .002; ΔSRMR = -.002), and RMSEA was slightly worse in the ESEM model (.072 vs. .067). The information criteria also favored the more parsimonious CFA specification, with lower AIC, BIC, and SABIC values for the ASB-anchored S-1 CFA model. These results indicated that the modest improvement in fit from freeing ESEM cross-loadings did not compensate for the reduction in degrees of freedom and corresponding increase in model complexity. In other words, after accounting for the ASB-anchored general HM factor, relatively little residual cross-domain covariance remained for ESEM cross-loadings to explain. Therefore, the

ASB-anchored S-1 CFA model was retained as the final measurement model. The retained ASB-anchored bifactor S-1 CFA model is presented in Figure X.

Figure 13

ASB-Anchored Bifactor S-1 CFA Model of Hostile Masculinity (S2)



Note. Values on single-headed paths are standardized factor loadings. Error paths are labeled with standardized residual variances. HM = hostile masculinity; ASB = adversarial sexual beliefs; HTW = hostility toward women; SOD = sexual dominance.

Internal Consistency Reliability

Internal consistency reliability was evaluated for the retained ASB-anchored bifactor S-1 CFA model, consisting of a general HM factor and three specific factors (i.e., HTW distrust, HTW resentment and SOD). McDonald's ω was computed to evaluate the overall reliability of the general HM factor and each specific factor (McDonald, 1999). Because the retained model included both a general factor and specific factors, omega hierarchical (ωH) and omega hierarchical subscale (ωHS) were also examined, as recommended reliability indices for evaluating bifactor models (Rodriguez et al., 2016). As a result, the general HM factor demonstrated excellent internal consistency, with an overall omega of .960 and an omega hierarchical of .872. These estimates indicated that the overall HM score was highly reliable and that the majority of its reliable variance was attributable to the general HM dimension rather than the specific factors. This finding is consistent with the bifactor S-1 results, which suggested a strong general HM construct underlying the retained indicators.

Reliability estimates for the specific factors also revealed a differentiated pattern. The HTW distrust-specific factor showed good overall reliability (ω = .871), but little reliable variance remained unique to the specific factor after accounting for the general HM factor (ωHS = .195). The HTW resentment-specific factor showed a similar pattern, with good overall reliability (ω = .872) but only modest unique reliable variance (ωHS = .271). In contrast, the SOD-specific factor showed good overall reliability (ω = .875) and retained substantially more unique reliable variance (ωHS = .541). Thus, the reliable variance in the HTW-specific factors was largely attributable to the general HM factor, whereas SOD retained comparatively greater construct-specific variance. See Table 27 for the reliability estimates.

Table 27

Reliability Estimates for the Retained ASB-Anchored Bifactor S-1 CFA Model

Score	Specific Factor	ω	ωH	ωHS
Full HM scale	—	0.96	0.872	—
HTW distrust specific	HTW_s1	0.871	0.676	0.195
HTW resentment specific	HTW_s2	0.872	0.602	0.271
SOD specific	SOD_s	0.875	0.334	0.541

Note. HM = hostile masculinity; HTW = hostility toward women; SOD = sexual dominance; ω = omega total; ωH = omega hierarchical; ωHS = omega hierarchical subscale. For the full HM scale, ωH represents the proportion of reliable variance attributable to the general HM factor. For specific factors, ωHS represents the proportion of reliable variance uniquely attributable to the specific factor after accounting for the general factor.

Overall, the omega-based reliability estimates further supported the reliability of the ASB-anchored bifactor S-1 CFA model. The results were consistent with the previous findings from factor loadings, ECV and PUC, indicating that ASB functions as the core indicator of the general HM construct, whereas the HTW distrust and resentment factors were largely explained by the broader HM construct and the SOD factor retained comparatively greater construct-specific variance.

Discussion

Study 2 examined the dimensional structure of HM by comparing higher-order, bifactor, and bifactor S-1 specifications in both CFA and ESEM frameworks. Extending the item-level findings from Study 1, Study 2 provided further support for the two-factor structure of HTW and evaluated how ASB, SOD, HTW distrust, and HTW resentment functioned together within broader measurement models of HM.

Overall, the results clearly evidenced the multidimensionality of HM and through the correlated CFA and ESEM models, there were meaningful associations among ASB, HTW, and SOD domains, indicating shared variance across hostile masculinity indicators. Although the higher-order model fit adequately, it did not separate general from domain-specific variance, whereas the conventional bifactor models were difficult to interpret without a clear reference domain. These findings supported the use of S-1 models, which allowed the general HM factor to be defined through a substantively meaningful anchor while estimating residual domain-specific variance.

Across the S-1 models, ASB emerged as a particularly coherent anchor for the general HM factor. Although the HTW resentment-anchored S-1 ESEM model showed the strongest global fit, the ASB-anchored S-1 models produced a stable and interpretable general factor across CFA and ESEM frameworks. The ASB-anchored S-1 CFA model was ultimately retained because it provided a favorable balance of fit, parsimony and interpretability. Theoretically, this suggests that adversarial and dominance-oriented beliefs about heterosexual relationships (i.e., ASB), including male control, female manipulation, and gender antagonism, capture a central component of HM, whereas power- and control-oriented sexual motives (i.e., SOD) and distrust- or resentment-based hostility toward women (i.e., HTW) retain meaningful residual domain-specific variance. Reliability analyses further supported this interpretation, showing excellent reliability for the general HM factor and comparatively greater unique reliable variance for SOD than for the HTW-specific factors.

Meanwhile, internal measurement structure alone might not be sufficient to determine which S-1 specification provides the most valid representation of HM. Because HM is theorized as a proximal risk factor for sexual aggression perpetration, model selection should also consider whether the resulting factor scores show theoretically coherent associations with sexual aggression and established Confluence Model risk factors. Study 3 therefore examined the construct and incremental criterion validity of the leading S-1 specifications, particularly the HTW resentment-anchored and ASB-anchored models, by testing which specification yields the most theoretically coherent pattern of associations between HM-related latent factors, sexual aggression, and other Confluence Model risk factors.

CHAPTER FOUR

Study 3 (S3)

Method

Participants

A total of 350 men were recruited through CloudResearch. After excluding 21 participants who failed the reCAPTCHA/data-quality screening, the final analytic sample consisted of 329 participants. Participants were all heterosexual men living in the United States, ranging in age from 18 to 80 years (*M* = 41.46, *SD* = 12.38). In terms of race/ethnicity, most participants identified as European American/White (*n* = 252, 76.6%), followed by African American/Black (*n* = 34, 10.3%), Asian American (*n* = 31, 9.4%), Hispanic/Latino(a) (*n* = 28, 8.5%), American Indian/Alaska Native (*n* = 4, 1.2%), Other (*n* = 3, 0.9%), Native Hawaiian/Pacific Islander (*n* = 1, 0.3%), and Middle Eastern/Arabic (*n* = 1, 0.3%). Regarding education, nearly half of participants reported having a bachelor's degree (*n* = 163, 49.5%), followed by a graduate degree (*n* = 57, 17.3%), some college (*n* = 41, 12.5%), a high school diploma or GED (*n* = 37, 11.2%), and an associate degree (*n* = 30, 9.1%); one participant preferred not to report education level (*n* = 1, 0.3%). Annual household income was broadly distributed, with 10.0% reporting less than $25,000 (*n* = 33), 14.3% reporting $25,000-$49,999 (*n* = 47), 20.7% reporting $50,000-$74,999 (*n* = 68), 18.2% reporting $75,000-$99,999 (*n* = 60), 21.9% reporting $100,000-$149,999 (*n* = 72), and 14.0% reporting $150,000 or more (*n* = 46); three participants preferred not to report income (*n* = 3, 0.9%). Demographic characteristics of the Study 3 sample are presented in Table 28.

Table 28

Demographic Characteristics of Study 3

Characteristic		**n**	**%**
Race/ethnicity	American Indian/Alaska Native	4	1.2
	Asian American	31	9.4
	Native Hawaiian/Pacific Islander	1	0.3
	African American/Black	34	10.3
	European American/White	252	76.6
	Middle Eastern/Arabic	1	0.3
	Hispanic/Latino(a)	28	8.5
	Other	3	0.9
Education	Some high school or less	0	0
	High school diploma/GED	37	11.2
	Some college	41	12.5
	Associate degree	30	9.1
	Bachelor's degree	163	49.5
	Graduate degree	57	17.3
Annual household income	< $25,000	33	10
	$25,000-$49,999	47	14.3
	$50,000-$74,999	68	20.7
	$75,000-$99,999	60	18.2
	$100,000-$149,999	72	21.9
	$150,000+	46	14

Note. N = 329. Percentages are based on the full Study 3 sample. Race/ethnicity categories were not mutually exclusive; therefore, percentages may sum to more than 100%. One participant did not report education level, and three participants did not report annual household income.

Procedure

A total of 350 men living in the United States were recruited through CloudResearch, an online participant recruitment platform. Prior research suggests that CloudResearch yields higher-quality data than some other online recruitment platforms, including MTurk, Qualtrics Panels, and SONA (Douglas et al., 2023). CloudResearch prescreening criteria were used to recruit participants who appeared to meet the study inclusion criteria: participants were required to live in the United States, self-identify as heterosexual men, and be at least 18 years old. Eligible participants were directed to a Qualtrics survey, where responses were screened for automated or low-quality responses using Qualtrics-generated reCAPTCHA scores. Scores below .50 were flagged as indicating an increased likelihood of bot activity. Twenty-one participants were excluded for failing the reCAPTCHA screening, resulting in a final analytic sample of 329 participants. Participants were then directed to review an informed consent document and indicate their consent electronically before proceeding. After consenting, participants then completed a short screening survey first to reconfirm they meet the inclusion criteria. Participants who passed the rescreening procedure received access to the main survey, which included demographic questions and measures of HM and several Confluence Model-related constructs. Every participant was given 48 hours to complete the study. After completing the survey, participants were directed to a resource page that provided contact information for the research team and support resources related to sexual assault, mental health, crisis support, and emergency assistance.

Measures

Concurrent measures. In addition to completing demographic questions and the HM scales, participants filled out six questionnaires that have been widely used to measure the Confluence Model risk and outcome factors.

Impersonal Sex (IS). The Revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008) was used to assess IS, which encompasses attitudes and behaviors related to casual sexual relationships. The SOI-R is a self-report measure with 3 subscales: Behavior, Attitude, and Desire. The Behavior subscale used numerical responses (from 0 to 8 or more). A sample item reads, "With how many different partners have you had sexual intercourse on one and only one occasion?" The Attitude subscale included 3 items that employ a 5-point Likert scale with responses ranging from "totally disagree" (1) to "totally agree" (5). A sample item reads, "I do not need to be committed to a person to have sex with them." The Desire subscale consisted of 3 items, using a 5-point Likert scale ranging from "never" (1) to "nearly every day" (9). A sample item reads, "How often do you have fantasies about having sex with someone you are not in a committed romantic relationship with?" The measure has demonstrated robust internal consistency reliability in previous validation studies (Cronbach's $\alpha = .86$; Penke & Asendorpf, 2008).

Rape Myth Acceptance. The Illinois Rape Myth Acceptance Scale-Short Form (IRMA-SF; Payne et al., 1999) was adopted to assess rape myth acceptance. The IRMA-SF includes 20 items, including 3 negatively worded filler items that were used in scale calculation. The scale uses a 7-point Likert scale from 1 (*Strongly Disagree*) to 7

(*Strongly Agree)* and includes the following subscales: she asked for it (e.g., "If a woman is raped while she is drunk, she is at least somewhat responsible for letting things get out of control"), it wasn't really rape (e.g., "If a woman doesn't physically fight back, you can't really say that it was rape"), he didn't mean to (e.g., "When men rape, it is because of their strong desire for sex"), she wanted it (e.g., "Although most women wouldn't admit it, they generally find being physically forced into sex a real 'turn-on'"), she lied (e.g., "Women who are caught having an illicit affair sometimes claim that it was rape"), rape is a trivial event (e.g., "If a woman is willing to 'make out' with a guy, then it's no big deal if he goes a little further and has sex"), and rape is a deviant event (e.g., "Rape mainly occurs on the 'bad' side of town"). All the subscale scores, except filler items, were summed to form a total score, with higher scores representing a higher level of rape myth acceptance. The scale has been widely assessed and validated with high internal consistency (α = .87- .97; Barnett et al., 2018; Canan et al., 2023; Owens et al., 2021)

Acceptance of Interpersonal Violence. The Acceptance of Interpersonal Violence Scale (AIV; Burt, 1980) was applied. The scale includes 6 items and is scored on a 7-point Likert-type scale (1 = *Strongly Disagree* to 7 = S*trongly Agree*). A sample item includes Item 2, "Being roughed up is sexually stimulating to many women." An overall score was computed by calculating the mean of all items, with higher scores representing higher levels of AIV. The scale has been largely adopted in previous Confluence Model literature (Ray & Parkhill, 2023); though it tends to demonstrate low internal reliability (α = .58; Ogle et al., 2009)

Empathy. Empathy was measured using the 16-item Toronto Empathy Questionnaire (TEQ; Kourmousi et al., 2017), scored on a 5-point scale (0 = Never to 4 = Always). Sample items include “Other people’s misfortunes do not disturb me a great deal” and “When a friend starts to talk about his/her problems, I try to steer the conversation towards something else.” The TEQ has demonstrated good validity and reliability (α = .72 to .85) from previous research (Kourmousi et al., 2017; Spreng et al., 2009; Ursoniu et al., 2021).

Psychopathy. Psychopathy was assessed using the 30-item Self-Report Psychopathy III Scale (SRP-III; Williams et al., 2007). Prior research has established good construct, convergent, and divergent validity of the measure in clinical, nonclinical, forensic, nonforensic, and community samples (Mahmut et al., 2011; Seibert et al., 2011). Participants rated their agreement with each item on a scale from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). Sample items include “I think I could'"beat" a lie detector.” and “It’s amusing to see other people get tricked.” An overall score was computed by calculating the mean of all items, with higher scores representing higher levels of psychopathy.

Childhood Sexual Abuse. Childhood sexual abuse (CSA) was assessed using the 7-item Sexual Abuse Scale (Whitmire et al., 1999), adapted for use with men (Catania et al., 2008). Participants were asked to think back to experiences before age 14 and indicate whether someone older had engaged them in a range of sexual situations. Sample items include “Did anyone older ever show their genitals to you?” and “Did anyone older ever touch your genitals?” Items were rated on a 4-point scale ranging from 1 (no) to 4 (many times). An overall score was computed by summing the items, with higher scores indicating more frequent childhood sexual abuse experiences. Prior research has reported high internal consistency for the measure (α = .92; Morokoff et al., 2009), and the scoring method has demonstrated validity in previous research (Whitmire et al., 1999).

Alcohol- and Drug-Related Sexual Behavior (ALCSEX & DRGSEX). Alcohol- and drug-related sexual behavior were assessed using items adapted from prior research on situational substance use in dating and sexual contexts (Abbey et al., 1998; Parkhill & Abbey, 2008). Participants completed parallel items assessing how often they typically used alcohol or drugs when on a date or when having consensual sexual intercourse, as well as how intoxicated or high they typically were in those situations. Sample items include "How often do you typically drink alcohol when you are on a date?" and "How high are you typically when you have consensual sexual intercourse?" Frequency items ranged from 0 (never) to 6 (every time), and intoxication/highness items ranged from no substance use to very intoxicated or very high. "Choose not to respond" responses were treated as missing. Separate alcohol-related and drug-related scores were computed, with higher scores indicating greater substance use in dating and consensual sexual situations.

Criterion Measure. Sexual aggression served as the criterion outcome variable in the present study. Participants completed a self-report measure assessing their prior experiences of sexual aggression perpetration.

Sexual Aggression. Sexual aggression was measured using the "tactics first" version of the Sexual Experiences Survey – Short Form Perpetration (SES-SFP; Abbey et al., 2005; Koss et al., 1987, 2007). As the gold standard measure of self-report sexual aggression perpetration, the Sexual Experience Survey has demonstrated construct validity across extensive studies. Respondents were asked to indicate the frequency of their perpetration of each specific sexually aggressive behavior against women. Scoring followed the "combined outcomes separated severity ranking" and the "highest severity rank" method (See Davis et al., 2014 for review), which includes 35 items assessing outcomes including unwanted sexual contact, attempted rape, completed rape through verbal coercion (e.g., lies, continual pressure, guilt induction, or criticism), incapacitation (i.e., alcohol or other drug intoxication), or force (i.e., physical force or threat of harm). Sample items include "*Since the age of 14, have you ever told a woman lies or overwhelmed her with continual arguments and pressure in order to* make her remove some of her clothes?" and "*Since the age of 14, have you ever threatened to physically harm a woman or someone close to her in order to* put your penis, fingers, or objects into her vagina?" Participants' scores were classified based on the most severe act reported: 0 = no history of sexual aggression, 1 = forced sexual contact, 2 = verbal coercion, 3 = attempted rape, 4 = completed rape.

Statistical Analysis

Convergent and Nomological Validity. To fulfill the fourth objective of establishing evidence for the convergent and nomological validity of the revised HM factors, statistical techniques were utilized, including bivariate correlations and multiple regression analyses. An HM total score, calculated across all retained ASB, HTW, and SOD items, was included as a conventional benchmark for comparison with the retained S-1 factor scores.

Bivariate correlations. Initial exploratory analysis of convergent validity was conducted using factor scores derived from the retained measurement models identified in Study 2. Specifically, bivariate correlations were computed between the estimated HM factor scores and scores on several established instruments representing theoretically related constructs within the Confluence Model of sexual aggression. Pearson's product-moment correlation coefficients were computed in the *R* package 'psych' version 2.5.6 (Revelle & Revelle, 2015) and p-values were adjusted using a Holm correction (Holm, 1979). Correlation coefficients with absolute values of .00 to .29 were interpreted as slight/negligible correlations; correlations of .30 to .49 as low correlations; .50 to .69 as moderate correlations; and .7 or higher as high correlations (Hinkle et al., 2003).

***Multiple Regression*.** To further evaluate the nomological network of the revised HM factor scores, multiple regression models were used to examine the extent to which the revised HM general and specific factors accounted for variance in theoretically relevant external validation measures. These included concurrent measures of sociosexuality (SOI-R), rape myth acceptance (IRMA-SF), acceptance of interpersonal violence (AIV), empathy (TEQ), psychopathic traits (SRP-III), childhood sexual trauma (CSA), alcohol-related sexual behavior (ALCSEX), and drug-related sexual behavior (DRGSEX), and a criterion measure of sexual aggression (SES-SFP). Separate regression models were estimated for each external validation measure, with the revised HM factor scores entered as predictors. All regression models were cross-validated using the ten-fold approach to address potential overfitting: Participants were divided into 10 approximately equal folds; each model was trained on nine folds and evaluated on the remaining fold. The process was repeated until every observation had served once as held-out data. The same participants and fold assignments were used across competing models to support direct comparisons. Predictive performance was evaluated using cross-validated mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and predictive R^2 (Shmueli, 2010; Stone, 1974). Lower MSE, MAE, and RMSE values indicated better out-of-sample prediction, whereas higher predictive R^2 values indicated greater explained variation in held-out SES-SFP scores. MAE and RMSE were selected because MAE is less sensitive to large prediction errors, and RMSE more heavily penalizes large prediction errors (Chai & Draxler, 2014; Willmott & Matsuura, 2005).

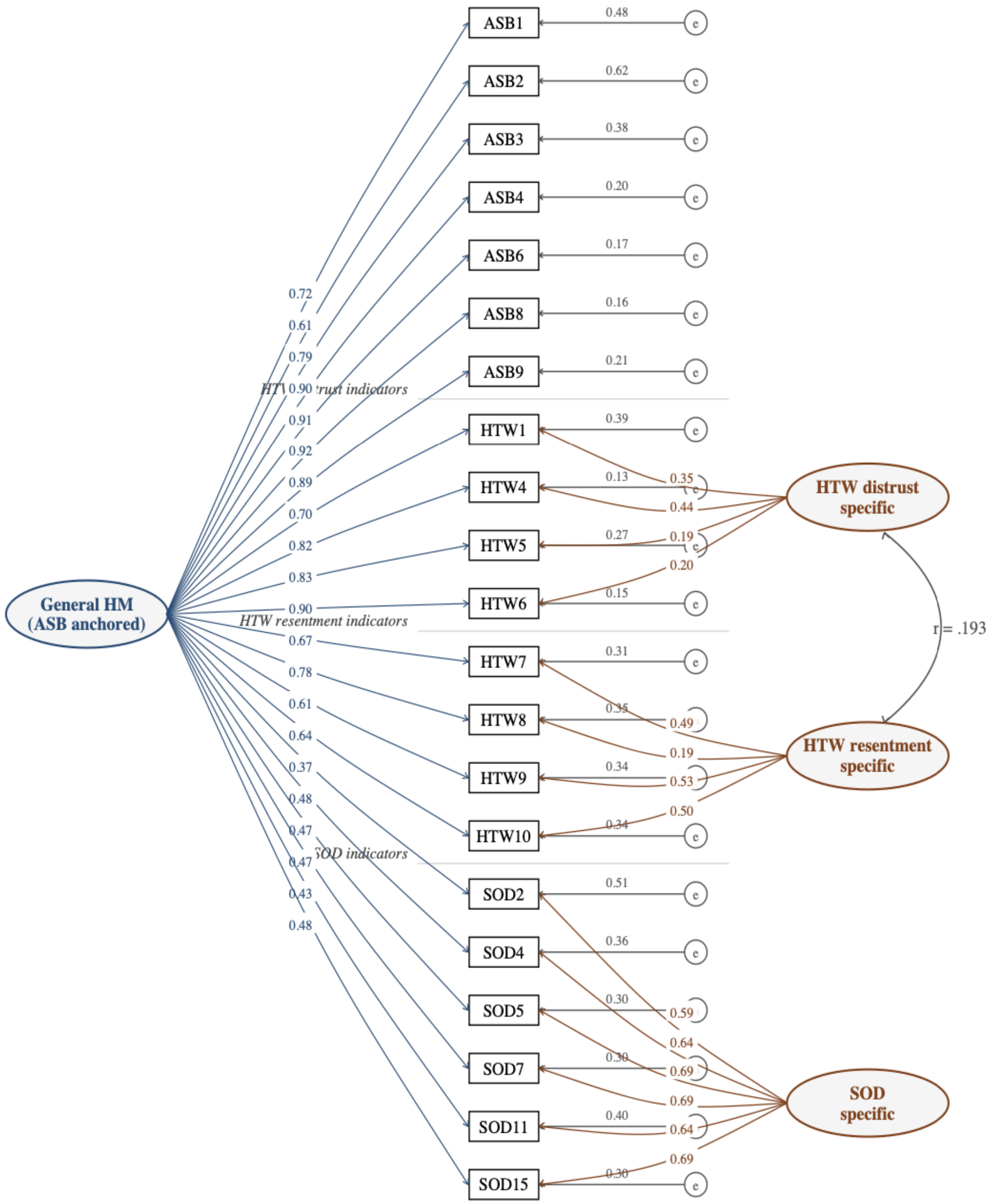
Results

Before evaluating external validity, the ASB-anchored S-1 CFA and the HTW resentment-anchored S-1 ESEM models were re-estimated using the Study 3 sample. As results, the ASB-anchored S-1 CFA model showed acceptable fit, $\chi^2(174) = 451.95$, $p < .001$, CFI = .942, TLI = .930, RMSEA = .079, 90% CI [.070, .088], SRMR = .046, AIC = 21,949.94, BIC = 22,246.03, and SABIC = 21,998.62 (see Figure 14). The ASB-anchored general HM factor was strongly defined by ASB and HTW items, whereas SOD items loaded more strongly on the residual SOD-specific factor. The HTW resentment-anchored S-1 ESEM model showed somewhat stronger global fit, $\chi^2(141) = 406.02$, $p < .001$, CFI = .955, TLI = .933, RMSEA = .080, 90% CI [.070, .091], SRMR = .030, AIC = 21,894.30, BIC = 22,315.66, and SABIC = 21,963.57. Information criteria provided mixed evidence: the HTW resentment-anchored model had lower AIC and

SABIC values, whereas the ASB-anchored model had a lower BIC. However, its residual specific-factor structure was less interpretable: the ASB-specific factor was defined by only a small subset of ASB items, and the exploratory residual factors did not align cleanly with their intended domains. See Table 29-30. The results generally replicated the patterns observed in Study 2. External validity analyses therefore proceeded using factor scores from both models, along with the HM total score (*M* = 2.89, *SD* = 1.25) as a conventional observed-score comparison.

Figure 14

ASB-anchored bifactor S-1 CFA model (S3)



Note. Values on single-headed paths are standardized factor loadings. Error paths are labeled with standardized residual variances. HM = hostile masculinity; ASB = adversarial sexual beliefs; HTW = hostility toward women; SOD = sexual dominance.

Table 29

Standardized Loadings for the HTW resentment-Anchored Bifactor S-1 ESEM

Item	HM_HTW resentment	ASB_s	SOD_s	HTW_F1_s
HTW7	**0.8**			
HTW8	**0.824**			
HTW9	**0.747**			
HTW10	**0.766**			
HTW1	**0.685**	-0.113	0.296	0.096
HTW4	**0.784**	-0.124	**0.397**	0.028
HTW5	**0.77**	-0.012	**0.362**	-0.012
HTW6	**0.84**	-0.047	**0.389**	-0.013
ASB1	**0.661**	**0.594**	-0.003	-0.001
ASB2	**0.593**	0.101	0.105	0.122
ASB3	**0.715**	0.52	0.077	0.009
ASB4	**0.765**	0.109	**0.401**	0.012
ASB6	**0.778**	0.028	**0.458**	-0.007
ASB8	**0.753**	0.008	**0.559**	-0.035
ASB9	**0.751**	0.042	**0.46**	-0.015
SOD2	**0.349**	0.026	-0.008	**0.599**
SOD4	**0.437**	0.119	-0.001	**0.626**
SOD5	**0.435**	0.022	0.026	**0.698**
SOD7	**0.491**	0.002	-0.074	**0.687**
SOD11	**0.429**	-0.026	-0.009	**0.66**
SOD15	**0.472**	-0.075	0.008	**0.715**

Latent Factor Correlations

Factor	HM	ASB_s	SOD_s	HTW distrust_s
HM	1			
ASB_s	0	1		
SOD_s	0	0.368	1	
HTW distrust_s	0	0.362	0.159	1

Note. HM_HTW resentment = general hostile masculinity factor anchored by HTW resentment items; ASB_s = residual Adversarial Sexual Beliefs-specific factor; SOD_s = residual Sexual Dominance-specific factor; HTW distrust_s = residual HTW distrust-specific factor. Loadings are standardized. The general factor was specified as orthogonal to all residual specific factors.

Table 30
Standardized Loadings for the ASB-Anchored Bifactor S-1 ESEM

Item	HM_ASB	HTW_s1	HTW_s2	SOD_s
ASB1	0.718			
ASB2	0.614			
ASB3	0.787			
ASB4	0.897			
ASB6	0.911			
ASB8	0.917			
ASB9	0.888			
HTW1	0.7	0.348		
HTW4	0.823	0.442		
HTW5	0.831	0.188		
HTW6	0.897	0.205		
HTW7	0.671		0.493	
HTW8	0.784		0.191	
HTW9	0.613		0.531	
HTW10	0.64		0.498	
SOD2	0.368			0.594
SOD4	0.477			0.643
SOD5	0.472			0.687
SOD7	0.465			0.695
SOD11	0.432			0.641
SOD15	0.476			0.685

Latent Factor Correlations

Factor	HM_ASB	HTW_s1	HTW_s2	SOD_s
HM_ASB	1			
HTW_s1	0	1		
HTW_s2	0	0.193	1	
SOD_s	0	0	0	1

Note. HM_HTW resentment = general hostile masculinity factor anchored by HTW resentment items; ASB_s = residual Adversarial Sexual Beliefs-specific factor; SOD_s = residual Sexual Dominance-specific factor; HTW distrust_s = residual HTW distrust-specific factor. Loadings are standardized. The general factor was specified as orthogonal to all residual specific factors.

Bivariate Correlations

Bivariate correlations provided evidence of convergent validity for the unidimensional HM total score and both S-1 specifications. The HM total demonstrated high positive correlations with rape myth acceptance (IRMA-SF; $r = .70$) and acceptance of interpersonal violence (AIV; $r = .69$), moderate positive correlations with psychopathic traits (SRP-SF; $r = .53$), and a low negative correlation with empathy (TEQ; $r = −.41$). It also showed positive correlations with childhood sexual abuse ($r = .36$), alcohol-related sexual behavior ($r = .26$), drug-related sexual behavior ($r = .34$), sociosexuality ($r = .25$), and sexual aggression ($r = .43$). For the retained ASB-anchored S-1 CFA model, the general HM factor demonstrated the expected pattern of associations with external validation measures, showing strong positive correlations with rape myth acceptance (IRMA-SF; $r = .64$) and acceptance of interpersonal violence (AIV; $r = .64$), a moderate positive correlation with psychopathic traits (SRP-SF; $r = .47$), a moderate negative correlation with empathy (TEQ; $r = −.39$), and smaller positive correlations with sociosexuality (SOI-R; r = .20) and sexual aggression (SES; $r = .35$). Among the residual specific factors, the HTW resentment-specific factor retained meaningful associations with psychopathic traits ($r = .33$), rape myth acceptance ($r = .33$), sexual aggression ($r = .38$), and empathy ($r = −.27$), whereas the SOD-specific factor remained positively associated with impersonal sex ($r = .22$), psychopathic traits ($r = .23$), acceptance of interpersonal violence ($r = .21$), and sexual aggression ($r = .22$). In contrast, the HTW distrust-specific factor exhibited few unique associations after accounting for the general HM factor. The HTW resentment-anchored S-1 ESEM model produced a similar pattern of correlations for the general HM factor, indicating comparable convergent validity. However, its residual specific factors demonstrated fewer factor loadings with fewer meaningful associations with the external validation measures, particularly the sexual aggression outcome, suggesting that less theoretically meaningful variance remained after extraction of the general HM factor. See Table 31 for the full set of bivariate correlations between the HM total score, S-1 factor scores, and external validation measures.

Table 31
Bivariate Correlations Between S-1 Factor Scores and External Validation Measures

External measure	HM_total	HTW resentment-Anchored S−1 ESEM General HM	HTW-F1 specific	ASB specific	SOD specific	ASB-Anchored S−1 CFA General HM	HTW specific 1	HTW specific 2	SOD specific
SOI-R total	.25***	.21**	-0.06	0.07	.21**	.20*	0.12	0.08	.22**
SRP-SF	.53***	.54***	0.09	0.02	.21**	.47***	0.06	.33***	.23**
TEQ	−.41***	-.46***	0.03	-0.03	0.02	-.39***	-.24**	-.27***	-0.01
AIV	.69***	.67***	.21**	0.16	.22**	.64***	0.08	.23**	.21**
IRMA-SF	.70***	.69***	.21**	0.14	0.18	.64***	0.03	.33***	0.17
CSA	.36***	.36***	0.01	-0.02	0.13	.30***	-0.08	.31***	0.16
ALCSEX	.26***	.21*	0.08	0.06	.21**	.20*	-0.05	0.13	.21*
DRGSEX	.34***	.41***	-0.13	-0.08	0.02	.31***	0.06	.37***	0.08
SES	.43***	.43***	0.04	-0.07	0.18	.35***	-0.08	.38***	.22**

Note. HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOD = sexual dominance; SOI-R = Sociosexual Orientation Inventory-Revised; SRP-SF = Self-Report Psychopathy-Short Form; TEQ = Toronto Empathy Questionnaire; AIV = Acceptance of Interpersonal Violence; IRMA-SF = Illinois Rape Myth Acceptance Scale-Short Form; CSA = childhood sexual abuse; ALCSEX = alcohol-related sexual behavior; DRGSEX = drug-related sexual behavior; SES = Sexual Experiences Survey. HTW-F1 and HTWF2 refer to the HTW distrust and HTW resentment, respectively. For the ASB-anchored S-1 CFA model, HTW Specific 1 and HTW Specific 2 refer to the residual HTW-specific factors. p < .05. p < .01. p < .001.

Multiple Regressions

To evaluate the nomological network of the revised HM scores, 10-fold cross-validated regression models compared the predictive performance of the unidimensional HM total, the ASB-anchored S-1 factor scores, and the HTW resentment-anchored S-1 factor scores across the external validation measures (See Table 32). The ASB-anchored model demonstrated the strongest prediction for seven of the nine outcomes, including rape myth acceptance, CV-R^2 = .504; psychopathic traits, CV-R^2 = .332; sexual aggression severity-frequency, CV-R^2 = .283; empathy, CV-R^2 = .204; drug-related sexual behavior, CV-R^2 = .186; childhood sexual abuse, CV-R^2 = .174; and alcohol-related sexual behavior, CV-R^2 = .066. The HTW resentment-anchored model performed best for sociosexuality, CV-R^2 = .088, whereas the unidimensional HM total performed best for acceptance of interpersonal violence, CV-R^2 = .468. Most notably, the ASB-anchored model explained more out-of-sample variance in the outcome, sexual aggression, than either the HTW resentment-anchored model or the unidimensional HM total (CV-R^2 = .283, .205, and .168, respectively). Within the ASB-anchored model, the standardized regression coefficients further indicated that general HM was consistently associated with all external measures (See Table 33). HTW resentment-specific variance showed additional associations with rape myth acceptance, acceptance of interpersonal violence, psychopathic traits, childhood sexual abuse, drug-related sexual behavior, and sexual aggression, whereas SOD-specific variance uniquely predicted several outcomes, including sociosexuality, rape myth acceptance, psychopathic traits, alcohol-related sexual behavior, and sexual aggression. In contrast, HTW distrust-specific variance showed fewer unique associations and was negatively associated with empathy, childhood sexual abuse, and sexual aggression (see Table 34).

Table 32
Ten-Fold Cross-Validated Predictive Performance of the HM-total and S-1 Factor-Score Models

Outcome	N	HM-total CV-R^2	ASB-anchored S-1 CV-R^2	HTW resentment-anchored S-1 CV-R^2
SOI-R total	323	0.055	0.069	0.088
IRMA-SF	328	0.472	0.504	0.478
AIV	327	0.468	0.455	0.462
TEQ	328	0.152	0.204	0.175
SRP-SF	328	0.282	0.332	0.305
CSA	327	0.104	0.174	0.104
ALCSEX	327	0.058	0.066	0.05
DRGSEX	326	0.09	0.186	0.157
SES-SFP	329	0.168	0.283	0.205

Note. CV-R^2 = 10-fold cross-validated coefficient of determination; HM = hostile masculinity; HTW = hostility toward women; ASB = adversarial sexual beliefs; SOI-R = Sociosexual Orientation Inventory–Revised; IRMA-SF = Illinois Rape Myth Acceptance Scale–Short Form; AIV = Acceptance of Interpersonal Violence; TEQ = Toronto Empathy Questionnaire; SRP-SF = Self-Report Psychopathy–Short Form; CSA = childhood sexual abuse; ALCSEX = alcohol-related sexual behavior; DRGSEX = drug-related sexual behavior; SES-SFP = Sexual Experiences Survey–Short Form Perpetration. Bold values indicate the strongest predictive performance for each outcome.

Table 33

Standardized Regression Coefficients for HTWF2-Anchored S-1 Factor Scores Predicting External Validation Measures

Outcome	N	General HM	ASB Specific	SOD Specific	HTW Distrust Specific
SOI-R	323	.201*** (.053)	-.202*** (.061)	.083 (.057)	.263*** (.057)
IRMA-SF	328	.672*** (.040)	.117* (.045)	-.023 (.042)	.091* (.043)
AIV	327	.587*** (.044)	.151** (.050)	-.069 (.047)	.088 (.047)
TEQ	328	-.468*** (.050)	.063 (.057)	.021 (.053)	.025 (.054)
SRP-III	328	.542*** (.046)	.001 (.053)	-.089 (.049)	.182*** (.050)
CSA	327	.370*** (.052)	-.042 (.059)	-.078 (.055)	.131* (.056)
ALCSEX	327	.195*** (.054)	-.021 (.062)	.018 (.058)	.205*** (.058)
DRGSEX	326	.434*** (.050)	-.152** (.057)	-.100 (.054)	.063 (.054)
SES-SFP	329	.449*** (.049)	-.023 (.056)	-.148** (.052)	.175** (.053)

Note. Values are standardized regression coefficients, with standard errors in parentheses. HM = hostile masculinity; ASB = adversarial sexual beliefs; SOD = sexual dominance; HTW = hostility toward women. $p < .05$. $p < .01$. $p < .001$.

Table 34

Standardized Regression Coefficients for ASB-Anchored S-1 Factor Scores Predicting External Validation Measures

Outcome	N	General HM	HTW Distrust specific	HTW Resentment Specific	SOD Specific
SOI-R	323	.183*** (.053)	.111* (.054)	.029 (.054)	.211*** (.053)
IRMA-SF	328	.630*** (.038)	-.063 (.039)	.302*** (.039)	.132*** (.039)
AIV	327	.539*** (.044)	-.044 (.045)	.236*** (.045)	.144** (.044)
TEQ	328	-.364*** (.049)	-.169*** (.050)	-.220*** (.050)	.015 (.049)
SRP-III	328	.451*** (.045)	-.024 (.046)	.298*** (.046)	.199*** (.045)
CSA	327	.299*** (.049)	-.155** (.050)	.315*** (.050)	.120* (.049)
ALCSEX	327	.201*** (.053)	-.082 (.054)	.118* (.054)	.189*** (.053)
DRGSEX	326	.295*** (.049)	-.032 (.050)	.361*** (.050)	.039 (.049)
SES-SFP	329	.343*** (.046)	-.173*** (.047)	.383*** (.047)	.175*** (.046)

Note. Values are standardized regression coefficients, with standard errors in parentheses. HM = hostile masculinity; HTW = hostility toward women; SOD = sexual dominance; SOI-R = Sociosexual Orientation Inventory-Revised; IRMA-SF = Illinois Rape Myth Acceptance Scale-Short Form; AIV = Acceptance of Interpersonal Violence; TEQ = Toronto Empathy Questionnaire; SRP-III = Self-Report Psychopathy Scale; CSA = childhood sexual trauma; ALCSEX = alcohol-related sexual behavior; DRGSEX = drug-related sexual behavior; SES-SFP = Sexual Experiences Survey-Short Form Perpetration. $p < .05$. $p < .01$. $p < .001$.

Discussion

Study 3 provided evidence for the external validity of the revised HM factor structure in an independent sample of heterosexual men. Overall, the retained measurement models replicated the general pattern observed in Study 2. Although both ASB-anchored S-1 CFA and HTW resentment-anchored S-1 ESEM models demonstrated acceptable fit, the ASB-anchored solution produced a more theoretically interpretable factor structure. In particular, the general HM factor was strongly defined by ASB and HTW items, whereas the SOD items retained meaningful residual specificity. By contrast, the HTW resentment-anchored ESEM solution showed somewhat stronger global fit but yielded residual factors that were less conceptually coherent.

The pattern of associations supported the nomological validity of the ASB-anchored S-1 CFA model. The general HM factor was strongly associated with rape myth acceptance and acceptance of interpersonal violence, moderately associated with psychopathic traits and lower empathy, and more modestly associated with impersonal sex and sexual aggression history. This pattern was consistent with prior Confluence Model research, in which rape myth acceptance and acceptance of interpersonal violence have frequently been used as indicators or correlates of hostile masculinity (Ray & Parkhill, 2023). And the observed associations with impersonal sex, psychopathic traits, empathy, and sexual aggression align with the broader nomological network of the Confluence Model, in which hostile masculinity and impersonal sex represent the two primary proximal risk factors for sexual aggression, while psychopathy and empathy have been identified as related risk factors (Abbey et al., 2006; Malamuth et al., 1991, 1995; Mouilso & Calhoun, 2013; Tharp et al., 2013).

The cross-validated regression results provided additional support for the ASB-anchored specification. The ASB-anchored S-1 factor scores demonstrated the strongest predictive performance for seven of the nine external measures and accounted for more out-of-sample variance in sexual aggression than either the HTW resentment-anchored model or the unidimensional HM total mean score. The ASB-anchored model also outperformed the total mean for rape myth acceptance, psychopathic traits, empathy, childhood sexual abuse, alcohol- and drug-related sexual behavior, and sexual aggression. However, the unidimensional HM total performed best for acceptance of interpersonal violence, and the HTW resentment-anchored model performed best for impersonal sex. Thus, although the unidimensional HM total remained a useful and parsimonious indicator of overall HM, partitioning HM into general and domain-specific components generally provided greater predictive utility across the broader nomological network.

At the factor level, the ASB-anchored general HM factor and HTW resentment-specific factor showed the most consistent associations with the external validation measures, followed by the SOD-specific factor. The HTW resentment-specific factor retained unique associations with rape myth acceptance, acceptance of interpersonal violence, psychopathic traits, childhood sexual abuse, drug-related sexual behavior, and sexual aggression, suggesting that resentment-based hostility captured meaningful risk-related variance beyond general HM. The SOD-specific factor also showed unique associations with impersonal sex, rape myth acceptance, psychopathic traits, alcohol-related sexual behavior, and sexual aggression. In contrast, the HTW distrust-specific factor was less consistently associated with the broader risk-factor network and was negatively associated with sexual aggression. In the HTW resentment-anchored model, the general HM factor carried most of the associations with the external validation measures, whereas the residual ASB, SOD, and HTW-specific factors showed less consistent and less conceptually differentiated patterns of association.

CHAPTER FIVE

General Discussion

The present research aims to advance the measurement of hostile masculinity (HM) by evaluating HM at both the item and latent-construct levels. Guided by the view that HM is inherently multidimensional, the present research addressed a key limitation of prior work: the frequent reliance on unidimensional models or item parceling, both of which may obscure meaningful distinctions among HM subdimensions. Across three coordinated studies, the findings supported a multidimensional measurement model of HM, characterized by a strong general factor most coherently anchored in adversarial sexual beliefs (ASB), alongside residual specific dimensions reflecting sexual dominance (SOD) and two distinguishable forms of hostility toward women (HTW): distrust-based and resentment-based hostility.

Findings of Study 1 provided the critical item-level foundation for the research by evaluating the psychometric functioning of the existing HTW, ASB, and SOD items across two archived datasets. Results generally supported the adequacy of the item pools, but also identified several items with weaker or less stable psychometric performance. Across dimensionality checks, GRM-IRT models, item information, and item-content review, several reverse-worded items were repeatedly flagged and were therefore removed. DIF analyses were used to evaluate between-study measurement equivalence and did not serve as the sole basis for item removal. These findings are consistent with prior work suggesting that reverse-worded items can introduce respondent confusion, increase measurement error, and produce method effects. Removing these items improved the interpretability of the revised HM item pool and reduced the likelihood that later latent structure analyses would be distorted by poorly functioning indicators.

Study 1 also clarified the dimensional structure of the HM components, particularly HTW. However, prior Confluence Model research has included HTW as an important component of HM; relatively little attention has been given to the multidimensionality of HTW. The present findings suggested that HTW included two related but distinguishable dimensions. One dimension reflected distrust-based hostility toward women, including beliefs that women are deceptive, manipulative, untrustworthy, or likely to exploit men. The other dimension reflected resentment-based hostility, including anger, grievance, irritation, and blame toward women. This distinction is conceptually meaningful because distrust-based hostility may reflect suspicion and defensiveness, whereas resentment-based hostility more directly captures affective antagonism toward women. The dimensions may therefore have different implications for sexual aggression perpetration, a pattern that was later reflected in the Study 3 external validity findings.

This distinction shaped the Study 2 measurement models by requiring HM to be evaluated as a structure composed of ASB, SOD, HTW distrust, and HTW resentment. Study 2 extended these findings by evaluating how the retained ASB, SOD, HTW distrust, and HTW resentment items functioned together across CFA and ESEM frameworks, with higher-order, bifactor, and bifactor S-1 specifications estimated within each framework. The results provided clear evidence that HM is multidimensional. A unidimensional model fitted poorly, whereas models distinguishing ASB, SOD, and the two HTW dimensions fitted substantially better. However, conventional higher-order and symmetrical bifactor models were either less interpretable or showed estimation instability, suggesting that freely estimating a broad general HM factor without a clear reference domain did not provide the most useful representation of the construct. In contrast, bifactor S-1 models provided a stronger conceptual framework because they allowed the general HM factor to be anchored in a theoretically meaningful domain while estimating residual domain-specific variance.

Across the S-1 models, ASB emerged as the most coherent anchor for the general HM factor. Although the HTW resentment-anchored ESEM model showed strong global fit, the ASB-anchored S-1 CFA model provided the best balance of fit, parsimony, stability, and interpretability. In this model, ASB items strongly defined the general HM factor, HTW items also contributed meaningfully to the general factor, and SOD retained the most reliable residual specificity. Reliability indices further supported this interpretation: the general HM factor demonstrated excellent reliability, whereas the SOD-specific factor retained more unique reliable variance than the HTW-specific factors after accounting for the general HM factor. These findings suggest that ASB captures a central shared core of HM, whereas SOD represents a more distinct residual component.

Finally, Study 3 evaluated whether the retained measurement structure generalized to an independent sample and whether the resulting factor scores showed theoretically meaningful associations with external validation measures. The ASB-anchored S-1 CFA model replicated well in the Study 3 sample and again yielded a more interpretable factor structure than the HTW resentment-anchored ESEM alternative.

Bivariate correlations supported the nomological validity of the ASB-anchored model. The general HM factor was strongly associated with rape myth acceptance and acceptance of interpersonal violence, moderately associated with psychopathic traits and lower empathy, and positively associated with sociosexuality and sexual aggression. Residual specific factors showed more differentiated patterns: HTW resentment and SOD retained meaningful associations with several risk-relevant constructs, whereas HTW distrust showed fewer consistent associations and was less clearly aligned with sexual aggression risk.

The cross-validated regression models provided further support for this interpretation. The ASB-anchored S-1 factor scores produced the strongest predictive performance for seven of the nine external measures and accounted for more out-of-sample variance in sexual aggression than either the HTW resentment-anchored model or the unidimensional HM score. Although the total mean remained a useful and parsimonious indicator and performed best for acceptance of interpersonal violence, partitioning HM into general and domain-specific components generally improved predictive performance across the broader Confluence Model risk network. These findings should not be interpreted as evidence of causal prediction, but they suggest that the ASB-anchored model more effectively captures the shared and domain-specific aspects of HM most relevant to sexual aggression and related risk factors.

Implications

The present findings have several implications for HM measurement and Confluence Model research. First, they suggest that treating HM as a simple total score or parcel-based latent variable may obscure meaningful multidimensionality. Although a reliable general HM factor was identified, the residual SOD and HTW dimensions were not psychometrically interchangeable. In particular, SOD retained meaningful unique variance after accounting for general HM, suggesting that researchers may lose important information if SOD is absorbed entirely into a global HM composite.

Second, the findings support the use of bifactor S-1 modeling as a useful framework for HM measurement. Rather than estimating a symmetrical bifactor model in which all domains define the general factor equally, the S-1 approach requires the researcher to identify a theoretically meaningful anchor. In the present dissertation, ASB provided the clearest anchor for the general HM factor. This finding suggests that adversarial beliefs about heterosexual relationships, including beliefs about male entitlement, female manipulation, and gender antagonism, may represent a central organizing feature of HM as currently measured.

Third, the distinction between HTW distrust and HTW resentment may help refine theory. Distrust-based hostility appeared to capture suspicion or defensiveness toward women, whereas resentment-based hostility appeared more closely tied to anger and antagonism. The Study 3 findings suggested that resentment-based hostility was more consistently connected to the broader sexual aggression risk network than distrust-based hostility. This distinction may be useful for future Confluence Model research because not all negative beliefs about women appear equally relevant to sexual aggression risk.

Limitations and Future Directions

Several limitations of the current research should be considered. First, the samples were limited to men, and Study 3 focused specifically on heterosexual men in the United States. This sampling decision is consistent with the original Confluence Model's focus on male-perpetrated sexual aggression against women, but it limits generalizability. Future research should evaluate whether the revised HM structure operates similarly across more diverse gender identities, sexual orientations, racial/ethnic groups, cultural contexts, and age ranges. Measurement invariance testing will be especially important before comparing HM scores across groups. Second, although Study 1 used integrative data analysis (IDA) to combine two existing datasets, the contributing samples differed in recruitment context and demographic composition. DIF analyses suggested that noninvariance was localized rather than pervasive, but some item-level differences remained. Future studies should replicate the revised item pool in samples designed prospectively for HM measurement, rather than relying on secondary data collected for related but distinct research purposes. Third, reverse-coded items warrant additional attention. Across the measures examined in this dissertation, several reverse-coded items were among those flagged for problematic functioning and were therefore removed in the current study. However, the effects of reverse wording could not be separated from differences in item content. For example, in the ASB scale, the reverse-coded items focused on men's behavior, whereas many of the remaining items referred more directly to women. Future research should therefore examine how reverse-coded HM items can be constructed more effectively, including whether difficulties arise from wording direction, item complexity, or substantive content. Studies could also test whether reverse-coded items provide meaningful protection against acquiescent responding or whether comparable measurement quality can be achieved without them. Finally, the present dissertation focused primarily on measurement structure and nomological validity rather than testing the full Confluence Model as a structural model. Study 3 evaluated associations between HM factor scores and external validation measures, but it did not estimate a full latent SEM including HM, impersonal sex, distal risk factors, and sexual aggression simultaneously. Future research should integrate the ASB-anchored HM model into full Confluence Model tests to examine whether distinguishing general and specific HM components improves theoretical precision and explanatory power. The last limitation concerns the scope of the item pool. Although the present research focused on ASB, HTW, and SOD as core HM components, prior research has operationalized HM inconsistently, sometimes incorporating additional constructs such as rape myth acceptance and acceptance of interpersonal violence (Ray & Parkhill, 2023). Ray and Parkhill (2023) suggested that these latter constructs may be better conceptualized as indicators of attitudes supporting violence, a related but potentially distinct construct within the hostile masculinity pathway. Future research should therefore test broader item pools drawn from prior HM studies to determine whether additional indicators improve the measurement of HM or whether constructs such as rape myth acceptance and acceptance of interpersonal violence should be modeled separately from HM to better predict sexual aggression in full Confluence Model analyses.

Conclusion

Across three studies, this research provides evidence that HM is best understood as a multidimensional construct with a strong ASB-anchored general factor and meaningful residual specific dimensions. Study 1 refined the item pool and identified problematic items, Study 2 established the ASB-anchored bifactor S-1 CFA model as the most parsimonious and interpretable measurement structure, and Study 3 supported the convergent and nomological validity of this model in an independent sample. Taken together, these findings offer a clearer framework for measuring HM within the Confluence Model and caution against relying on untested total scores, parcels, or overly broad composites. By distinguishing general HM from residual HTW and SOD dimensions, the present work provides a stronger foundation for future research on the measurement, theory, and prevention-relevant study of sexual aggression risk.

APPENDICES

Appendix A. Institutional Review Board Approval Letter

OAKLAND UNIVERSITY

Institutional Review Board

Date: April 17, 2026

Study #: IRB-FY2025-266

Study Title: Project HMPM

Submission Type: Modification

IRB Decision: Exempt

Research Team:
Wenqi Zheng
Michele Purdie
Charlie Dimock, Dylan Weber

The IRB has reviewed the Modification submission related to the above referenced study and determined that the changes listed below do not affect the exemption status of the study. The Modification submission is Approved and the study remains Exempt per federal regulations.

The approved modifications are the following:

- Revise the collection plan to a) remove the SONA recruitment and b) move the community sample recruitment to CloudResearch instead of Prolific.
- Participants will receive $4.00 through CloudResearch for the completion of the Qualtrics survey that will take about 30 minutes.
- In the survey, CloudResearch ID will not be recorded.
- Removal of SONA and Prolific recruitment document and addition of CloudResearch recruitment document.
- Removal of SONA and Prolific Consent Forms and addition of CloudResearch Consent Form.
- Decrease the total number of participants from 700 to 350.

This submission includes the following approved documents:

- CouldResearch Recruitment document
- Date-Stamped CloudResearch Consent Form V 4/17/2026

The IRB approved (date-stamped) consent document(s) has been published under Attachments in the Submission Details page. Please make sure to use the most recent IRB approved version of the consent form in consenting participants.

You are approved to implement the aforementioned modifications. Please retain a copy of this notification for your records.

This letter can also be found in Cayuse under the Letters tab in the Submission Details page.

Appendix B. Sexual Dominance Scale

Sexual Dominance (SEXDOM)
(Nelson, 1979)

People have sexual relations for many reasons. We would like to know how important each of these reasons is to you. After considering each of the reasons listed below carefully, check the number that best indicates how important that reason is in your own sexual behavior.

I have sexual relations…

Variable Name	Variable Label	Response Options
SEXDOM1	Because it's the way I show that I really care about someone. **REVERSE CODE**	1 = not at all important 4 = neutral 7 = very important
SEXDOM2	Because I like the feeling that I have someone in my grasp.	
SEXDOM3	Because it makes me feel like someone cares about me. **REVERSE CODE**	
SEXDOM4	Because, like many people, I enjoy the conquest.	
SEXDOM5	Because it makes me feel masterful.	
SEXDOM6	Because it makes me feel as one with another person. **REVERSE CODE**	
SEXDOM7	Because I like the feeling of having another person submit to me.	
SEXDOM8	Because sex and love are as one to me. **REVERSE CODE**	
SEXDOM9	Because I like teaching less experienced people how to get off.	
SEXDOM10	Because it makes me feel intimate with my partner. **REVERSE CODE**	
SEXDOM11	Because in the act of sex more than at any other time,e I get the feeling that I can really influence how someone feels and behaves.	
SEXDOM12	Because of the feeling of closeness it brings to a relationship. **REVERSE CODE**	
SEXDOM13	Because I like it when my partner is really open and vulnerable to me. **REVERSE CODE**	
SEXDOM14	Because it's the way I show my partner I love her. **REVERSE CODE**	
SEXDOM15	Because when my partner finally surrenders to me, I get this incredibly satisfying feeling.	
SEXDOM16	Because I enjoy being affectionate and sharing my feelings. **REVERSE CODE**	

Appendix C. Hostility toward Women Scale

Hostility toward Women (HTW)
(Lonsway and Fitzgerald, 1995)

Please rate how strongly you agree or disagree with each of the following statements.

Variable Name	Variable Label	Response Options
HTW1	I feel that many times women flirt with men just to tease them or hurt them	1 = Strongly disagree 2 = Disagree 3 = Slightly disagree 4 = Neutral 5 = Slightly agree 6 = Agree 7 = Strongly agree
HTW2	I believe that most women tell the truth **REVERSE CODE**	
HTW3	I usually find myself agreeing with women **REVERSE CODE**	
HTW4	I think that most women would lie just to get ahead	
HTW5	Generally, it is safer not to trust women	
HTW6	When it really comes down to it, a lot of women are deceitful	
HTW7	I am easily angered by women	
HTW8	I am sure I get a raw deal from women in my life	
HTW9	Sometimes women bother me by just being around	
HTW10	Women are responsible for most of my troubles	

Appendix D. Adversarial Sexual Beliefs Scale

Adversarial Sexual Beliefs (ASB)
(Burt, 1980)

Please rate how strongly you agree or disagree with each of the following statements.

Variable Name	Variable Label	Response Options
ASB1	A woman will only respect a man who will lay down the law to her	1 = Strongly disagree 2 = Disagree 3 = Slightly disagree 4 = Neutral 5 = Slightly agree 6 = Agree 7 = Strongly agree
ASB2	Many women are so demanding sexually that a man just can't satisfy them	
ASB3	A man has got to show the woman who is boss right from the start, or he'll end up whipped.	
ASB4	Women are usually sweet until they've caught a man, but then they let their true self show.	
ASB5	A lot of men talk big, but when it comes down to it, they can't perform well sexually. **REVERSE CODE**	
ASB6	In a dating relationship, a woman is largely out to take advantage of a man.	
ASB7	Men are out for only one thing **REVERSE CODE**	
ASB8	Most women are sly and manipulative when they are out to attract a man	
ASB9	A lot of women seem to get pleasure in putting men down	

Appendix E. Acceptance of Interpersonal Violence Scale

Acceptance of Interpersonal Violence Scale (AIV)
(Burt, 1980)

Please rate how strongly you agree or disagree with each of the following statements.

Variable Name	**Variable Label**	**Value Labels / Anchors**
AIV1	People today should not use “an eye for an eye and a tooth for a tooth” as a rule for living.	1 = Strongly agree 2 = Agree 3 = Agree somewhat 4 = Neutral 5 = Disagree somewhat 6 = Disagree 7 = Strongly disagree
AIV2	Being roughed up is sexually stimulating to many women.	
AIV3	Many times, a woman will pretend she doesn’t want to have intercourse because she doesn’t want to seem loose, but she’s really hoping the man will force her.	
AIV4	A wife should move out of the house if her husband hits her. **REVERSE CODE**	
AIV5	Sometimes the only way a man can get a cold woman turned on is to use force.	
AIV6	A man is never justified in hitting his wife. **REVERSE CODE**	

Appendix F. Illinois Rape Myth Acceptance Scale - Short Form

Illinois Rape Myth Acceptance Scale - Short Form (IRMA-SF)
(Payne et al., 1999)

Variable Name	Variable Label	Response Options
SA-3	If a woman is raped while she is drunk, she is at least somewhat responsible for letting things get out of control	1 = Strongly disagree 2 = Disagree 3 = Slightly disagree 4 = Neutral 5 = Slightly agree 6 = Agree 7 = Strongly agree Choose not to respond
WI-5	Although most women wouldn't admit it, they generally find being physically forced into sex a real "turn-on"	
TE-5	If a woman is willing to "make out" with a guy, then it's no big deal if he goes a little further and has sex	
WI-1	Many women secretly desire to be raped	
FI-2	Most rapists are not caught by the police	
NR-1	If a woman doesn't physically fight back, you can't really say that it was rape.	
DE-2	Men from nice middle-class homes almost never rape	
LI-2	Rape accusations are often used as a way of getting back at men	
FI-3*	All women should have access to self-defense classes	
DE-3	It is usually only women who dress suggestively that are raped	
NR-3	If the rapist doesn't have a weapon, you really can't call it a rape	
DE-7	Rape is unlikely to happen in the woman's own familiar neighborhood	
TE-2	Women tend to exaggerate how much rape affects them	
LI-1	A lot of women lead a man on and then they cry rape	
FI-4*	It is preferable that a female police officer conduct the questioning when a woman reports a rape.	
SA-5	A woman who "teases" men deserves anything that might happen	
SA-8	When women are raped, it's often because the way they said "no" was ambiguous.	
MT-1	Men don't usually intend to force sex on a woman, but sometimes they get too sexually carried away.	
SA-1	A woman who dresses in skimpy clothes should not be surprised if a man tries to force her to have sex.	
MT-4	Rape happens when a man's sex drive gets out of control	

*FI=Filler Item. These ‘‘filler items’’ are included to minimize the potential for response bias and are therefore discarded before conducting any statistical analysis. They do not contribute to the total scale value.

Response options:
Respondents were asked to rate their level of agreement using a 7-point Likert scale ranging from 1 = "not at all agree" to 7 = "very much agree.”

Author changed response items to a 7-point scale from “Strongly Disagree” to “Strongly Agree”
Scores to the Filler Items are not counted towards the total scale value.

REFERENCES

Abbey, A., Helmers, B. R., Jilani, Z., McDaniel, M. C., & Benbouriche, M. (2021). Assessment of men's sexual aggression against women: An experimental comparison of three versions of the sexual experiences survey. *Psychology of Violence*, *11*(3), 253–263. https://doi.org/10.1037/vio0000378

Abbey, A., & Jacques-Tiura, A. J. (2011). Sexual assault perpetrators' tactics: Associations with their personal characteristics and aspects of the incident. *Journal of Interpersonal Violence*, *26*(14), 2866–2889. https://doi.org/10.1177/0886260510390955

Abbey, A., Jacques‐Tiura, A. J., & LeBreton, J. M. (2011). Risk factors for sexual aggression in young men: An expansion of the confluence model. *Aggressive Behavior*, *37*(5), 450–464. https://doi.org/10.1002/ab.20399

Abbey, A., McAuslan, P., & Ross, L. T. (1998). Sexual assault perpetration by college men: The role of alcohol, misperception of sexual intent, and sexual beliefs and experiences. *Journal of Social and Clinical Psychology; New York*, *17*(2), 167–195. http://search.proquest.com/docview/224853372/abstract/5B4CB7117ED34B33PQ/1

Abbey, A., Parkhill, M. R., BeShears, R., Clinton-Sherrod, A. M., & Zawacki, T. (2006). Cross-sectional predictors of sexual assault perpetration in a community sample of single African American and Caucasian men. *Aggressive Behavior*, *32*(1), 54–67. https://doi.org/10.1002/ab.20107

Abbey, A., Parkhill, M. R., & Koss, M. P. (2005). The effects of frame of reference on responses to questions about sexual assault victimization and perpetration. *Psychology of Women Quarterly*, *29*(4), 364–373. https://doi.org/10.1111/j.1471-6402.2005.00236.x

Alamer, A. (2022). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied examples. *Research Methods in Applied Linguistics*, *1*, 100005. https://doi.org/10.1016/j.rmal.2022.100005

Allsworth, J. E., Anand, M., Redding, C. A., & Peipert, J. F. (2009). Physical and sexual violence and incident sexually transmitted infections. *Journal of Women's Health*, *18*(4), 529–534. https://doi.org/10.1089/jwh.2007.0757

Anderson, C., & Anderson, K. (2008). Men who target women: Specificity of target, generality of aggressive behavior. *Aggressive Behavior*, *34*(6), Article 6. https://doi.org/10.1002/ab.20274

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Ayala, R. J. de. (2013). *The Theory and Practice of Item Response Theory*. Guilford Publications.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling: A*

Multidisciplinary Journal, *9*(1), 78–102. https://doi.org/10.1207/S15328007SEM0901_5

Bandalos, D. L. (2008). Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(2), 211–240. https://doi.org/10.1080/10705510801922340

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In *New Developments and Techniques in Structural Equation Modeling*. Psychology Press.

Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication*, *28*(3), 12–29. https://doi.org/10.1111/j.1460-2466.1978.tb01621.x

Barnett, M. D., Sligar, K. B., & Wang, C. D. C. (2018). Religious affiliation, religiosity, gender, and rape myth acceptance: Feminist theory and rape culture. *Journal of Interpersonal Violence*, *33*(8), 1219–1235. https://doi.org/10.1177/0886260516665110

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361–370. https://doi.org/10.1177/00131640021970592

Bauer, D. J., & Hussong, A. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*(2), 101–125. https://doi.org/10.1037/a0015583

Bird, E. R., Stappenbeck, C. A., Neilson, E. C., Gulati, N. K., George, W. H., Cooper, M. L., & Davis, K. C. (2019). Sexual victimization and sex-related drinking motives: How protective is emotion regulation? *The Journal of Sex Research*, *56*(2), 156–165. https://www.jstor.org/stable/26772834

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*(3), 201–213. https://www.jstor.org/stable/2488508

Brazil, K. J., Roy, S., Bubeleva, K. V., & Neumann, C. S. (2023). Sexual assault proclivity and sexual aggression in college men: Associations with psychopathic traits and sex drive. *Journal of Criminal Justice*, *87*, 102089. https://doi.org/10.1016/j.jcrimjus.2023.102089

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research, Second Edition*. Guilford Publications.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Brownmiller, S. (1993). *Against Our Will: Men, Women, and Rape*. Random House Publishing Group.

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology*, *38*(2), 217–230. https://doi.org/10.1037/0022-3514.38.2.217

Buss, D. M., & Schmitt, D. P. (1993). Sexual Strategies Theory: An evolutionary perspective on human mating. *Psychological Review*, *100*(2), 204–232. https://doi.org/10.1037/0033-295X.100.2.204

Calhoun, K. S., Bernat, J. A., Clum, G. A., & Frame, C. L. (1997). Sexual coercion and attraction to sexual aggression in a community sample of young men. *Journal of Interpersonal Violence*, *12*(3), 392–406. https://doi.org/10.1177/088626097012003005

Canan, S. N., Cozzolino, L., Myers, J. L., & Jozkowski, K. N. (2023). Does gender inclusive language affect psychometric properties of the Illinois Rape Myth Acceptance Scale–Short Form? A two-sample validation study. *Journal of Interpersonal Violence*, *38*(3–4), 3373–3394. https://doi.org/10.1177/08862605221106144

Casey, E., Masters, N., Beadnell, B., Hoppe, M., Morrison, D., & Wells, E. (2017). Predicting Sexual Assault Perpetration Among Heterosexually Active Young Men. *Violence Against Women*, *23*(1), Article 1. https://doi.org/10.1177/1077801216634467

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Check, J., & Malamuth, N. (1983). Sex-role stereotyping and reactions to depictions of stranger versus acquaintance rape. *Journal of Personality and Social Psychology*, *45*(2), Article 2. https://doi.org/10.1037/0022-3514.45.2.344

Choudhary, E., Smith, M., & Bossarte, R. M. (2012). Depression, anxiety, and symptom profiles among female and male victims of sexual violence. *American Journal of Men's Health*, *6*(1), 28–36. https://doi.org/10.1177/1557988311414045

Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research*, *40*(2), 235–259. https://doi.org/10.1207/s15327906mbr4002_4

Cohen, A. S., Kim, S.-H., & Wollack, J. A. (1996). An investigation of the likelihood ratio test for detection of differential item functioning. *Applied Psychological Measurement*, *20*(1), 15–26. https://doi.org/10.1177/014662169602000102

Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society*, *19*(6), 829–859. https://doi.org/10.1177/0891243205278639

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100. https://doi.org/10.1037/a0015914

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*(2), 365–380. https://doi.org/10.1037/0012-1649.44.2.365

Dai, S., Vo, T. T., Kehinde, O. J., He, H., Xue, Y., Demir, C., & Wang, X. (2021). Performance of polytomous IRT models with rating scale data: An investigation over sample size, instrument length, and missing data. *Frontiers in Education*, *6*, 721963. https://doi.org/10.3389/feduc.2021.721963

Davis, K. C., Gilmore, A. K., Stappenbeck, C. A., Balsan, M. J., George, W. H., & Norris, J. (2014). How to score the Sexual Experiences Survey? A comparison of nine methods. *Psychology of Violence*, *4*(4), 445–461. https://doi.org/10.1037/a0037494

Davis, K. C., & Logan-Greene, P. (2012). Young men's aggressive tactics to avoid condom use: A test of a theoretical model. *Social Work Research*, *36*(3), 223–231. https://doi.org/10.1093/swr/svs027

Dean, K. E., & Malamuth, N. M. (1997). Characteristics of men who aggress sexually and of men who imagine aggressing: Risk and moderating variables. *Journal of Personality and Social Psychology*, *72*(2), 449–455. https://doi.org/10.1037/0022-3514.72.2.449

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

Deitz, M., Williams, S., Rife, S., & Cantrell, P. (2015). Examining cultural, social, and self-related aspects of stigma in relation to sexual assault and trauma symptoms. *Violence Against Women*, *21*(5), Article 5. https://doi.org/10.1177/1077801215573330

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and

SONA. *PLOS ONE*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541–562. https://doi.org/10.1037/met0000083

Eid, M., Krumm, S., Koch, T., & Schulze, J. (2018). Bifactor models for predicting criteria by general and specific factors: Problems of nonidentifiability and alternative solutions. *Journal of Intelligence*, *6*(3), 42. https://doi.org/10.3390/jintelligence6030042

Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory for Psychologists*. Psychology Press. https://doi.org/10.4324/9781410605269

Eysenck, S. B. G., & Eysenck, H. J. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of personality description. *Psychological Reports*, *43*(3_suppl), 1247–1255. https://doi.org/10.2466/pr0.1978.43.3f.1247

Finkelhor, D., & Araji, S. (1986). Explanations of pedophilia: A four factor model. *The Journal of Sex Research*, *22*(2), 145–161. https://doi.org/10.1080/00224498609551297

Ford, J., MacCALLUM, R., & TAIT, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*, 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57–68. https://doi.org/10.1016/j.intell.2016.01.006

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, *37*(4), 827–838. https://doi.org/10.1177/001316447703700403

Groth, A. N., & Birnbaum, H. J. (2013). *Men Who Rape: The Psychology of the Offender*. Springer.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*(3), 233–256. https://doi.org/10.1177/109442819923002

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*(6), 723–744. https://doi.org/10.1177/014920639702300602

Holzinger, K. J., & Swineford, F. (n.d.). The Bi-factor method | Psychometrika. *Cambridge Core*. https://doi.org/10.1007/BF02287965

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International*

Journal of Human-Computer Interaction, *32*(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huntington, C., Pearlman, D. N., & Orchowski, L. (2022). The Confluence Model of sexual aggression: An application with adolescent males. *Journal of Interpersonal Violence*, *37*(1–2), 623–643. https://doi.org/10.1177/0886260520915550

Hurt, C. (1980). *Impact of Price Risk on Aggregate Farrowing Response* [Ph.D., University of Illinois at Urbana-Champaign]. https://www.proquest.com/docview/303021010/abstract/84268D4CB3F6467BPQ/1

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*, *9*(1), 61–89. https://doi.org/10.1146/annurev-clinpsy-050212-185522

Johansson, F., Edlund, K., Sundgot-Borgen, J., Björklund, C., Côté, P., Onell, C., Sundberg, T., & Skillgate, E. (2024). Sexual harassment, sexual violence and subsequent depression and anxiety symptoms among Swedish university students: A cohort study. *Social Psychiatry and Psychiatric Epidemiology*, *59*(12), 2313–2322. https://doi.org/10.1007/s00127-024-02688-0

Kearns, M. C., D'Inverno, A. S., & Reidy, D. E. (2020). The association between gender inequality and sexual violence in the U.S.. *American Journal of Preventive Medicine*, *58*(1), 12–20. https://doi.org/10.1016/j.amepre.2019.08.035

Kennedy, A. C., & Prock, K. A. (2018). "I still feel like I am not normal": A review of the role of stigma and stigmatization among female survivors of child sexual abuse, sexual assault, and intimate partner violence. *Trauma, Violence, & Abuse*, *19*(5), 512–527. https://doi.org/10.1177/1524838016673601

Kieftenbeld, V., & Natesan, P. (2012). Recovery of graded response model parameters: A comparison of marginal maximum likelihood and Markov chain Monte Carlo estimation. *Applied Psychological Measurement*, *36*(5), 399–419. https://doi.org/10.1177/0146621612446170

Kirwan, M., Lanni, D., Warnke, A., Pickett, S., & Parkhill, M. (2019). Emotion regulation moderates the relationship between alcohol consumption and the perpetration of sexual aggression. *Violence Against Women*, *25*(9), Article 9. https://doi.org/10.1177/1077801218808396

Kirwan, M., Westemeier, O., Hammett, J. F., Stappenbeck, C. A., & Davis, K. C. (2025). Integrating the Confluence Model and I^3 Model to predict sexual assault perpetration intentions. *Psychology of Addictive Behaviors*. https://doi.org/10.1037/adb0001059

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and*

Psychological Measurement, *54*(3), 757–765. https://doi.org/10.1177/0013164494054003022

Koss, M. P. (1996). The measurement of rape victimization in crime surveys. *Criminal Justice and Behavior*, *23*(1), 55–69. https://doi.org/10.1177/0093854896023001005

Koss, M. P., Abbey, A., Campbell, R., Cook, S., Norris, J., Testa, M., Ullman, S., West, C., & White, J. (2007). Revising the SES: A collaborative process to improve assessment of sexual aggression and victimization. *Psychology of Women Quarterly*, *31*(4), 357–370. https://doi.org/10.1111/j.1471-6402.2007.00385.x

Koss, M. P., Anderson, R., Peterson, Z. D., Littleton, H., Abbey, A., Kowalski, R., Thompson, M., Canan, S., White, J., McCauley, H., Orchowski, L., Fedina, L., Lopez, E., & Allen, C. (2024). The Revised Sexual Experiences Survey Victimization Version (SES-V): Conceptualization, modifications, items and scoring. *The Journal of Sex Research*, *61*(6), 839–867. https://doi.org/10.1080/00224499.2024.2358407

Koss, M. P., & Gidycz, C. A. (1985). Sexual Experiences Survey: Reliability and validity. *Journal of Consulting and Clinical Psychology*, *53*(3), 422–423. https://doi.org/10.1037/0022-006X.53.3.422

Koss, M. P., & Oros, C. J. (1982). Sexual Experiences Survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology*, *50*(3), 455–457. https://doi.org/10.1037/0022-006X.50.3.455

Kourmousi, N., Amanaki, E., Tzavara, C., Merakou, K., Barbouni, A., & Koutras, V. (2017). The Toronto Empathy Questionnaire: Reliability and validity in a nationwide sample of Greek teachers. *Social Sciences*, *6*(2), 62. https://doi.org/10.3390/socsci6020062

Lalumière, M. L., Chalmers, L. J., Quinsey, V. L., & Seto, M. C. (1996). A test of the mate deprivation hypothesis of sexual coercion. *Ethology & Sociobiology*, *17*(5), 299–318. APA PsycInfo® (619050295; 1997-02661-002). https://doi.org/10.1016/S0162-3095(96)00076-3

Levant, R. F., Rankin, T. J., Williams, C. M., Hasan, N. T., & Smalley, K. B. (2010). Evaluation of the factor structure and construct validity of scores on the Male Role Norms Inventory—Revised (MRNI-R). *Psychology of Men & Masculinity*, *11*(1), 25–37. https://doi.org/10.1037/a0017637

Lewis, M. (2007). Stepwise versus hierarchical regression: Pros and cons. In *Online Submission*. https://eric.ed.gov/?id=ED534385

Lisak, D., & Roth, S. (1988). Motivational factors in nonincarcerated sexually aggressive men. *Journal of Personality and Social Psychology*, *55*(5), 795–802. https://doi.org/10.1037/0022-3514.55.5.795

Little, T., Cunningham, W., Shahar, G., & Widaman, K. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling A Multidisciplinary Journal*, *9*, 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Loh, C., & Gidycz, C. (2006). A prospective analysis of the relationship between childhood sexual victimization and perpetration of dating violence and sexual assault in adulthood. *Journal of Interpersonal Violence*, *21*(6), Article 6. https://doi.org/10.1177/0886260506287313

Lonsway, K. A., & Fitzgerald, L. F. (1995). Attitudinal antecedents of rape myth acceptance: A theoretical and empirical reexamination. *Journal of Personality and Social Psychology*, *68*(4), 704–711. https://doi.org/10.1037/0022-3514.68.4.704

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201–226. https://doi.org/10.1146/annurev.psych.51.1.201

Mahmut, M. K., Menictas, C., Stevenson, R. J., & Homewood, J. (2011). Validating the factor structure of the Self-Report Psychopathy Scale in a community sample. *Psychological Assessment*, *23*(3), 670–678. https://doi.org/10.1037/a0023090

Malamuth, N. M. (1986). Predictors of naturalistic sexual aggression. *Journal of Personality and Social Psychology*, *50*(5), 953–962. https://doi.org/10.1037/0022-3514.50.5.953

Malamuth, N. M., Addison, T., & Koss, M. (2000). Pornography and sexual aggression: Are there reliable effects and can we understand them? *Annual Review of Sex Research*, *11*(1), 26–91. https://doi.org/10.1080/10532528.2000.10559784

Malamuth, N. M., & Hald, G. M. (2016). The confluence mediational model of sexual aggression. In *The Wiley Handbook on the Theories, Assessment and Treatment of*

Sexual Offending (pp. 53–71). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118574003.wattso003

Malamuth, N. M., Hald, G. M., & Koss, M. (2012). Pornography, individual differences in risk, and men's acceptance of violence against women in a representative sample. *Sex Roles*, *66*(7–8), 427–439. https://doi.org/10.1007/s11199-011-0082-6

Malamuth, N. M., Heavey, C. L., & Linz, D. (1996). The Confluence Model of sexual aggression: Combining hostile masculinity and impersonal sex.. *Journal of Offender Rehabilitation*, *23*(3–4), 13–37. https://doi.org/10.1300/J076v23n03_03

Malamuth, N. M., Lamade, R. V., Koss, M. P., Lopez, E., Seaman, C., & Prentky, R. (2021). Factors predictive of sexual violence: Testing the four pillars of the Confluence Model in a large diverse sample of college men. *Aggressive Behavior*, *47*(4), 405–420. https://doi.org/10.1002/ab.21960

Malamuth, N. M., Linz, D., Heavey, C. L., Bames, G., & Acker, M. (1995). Using the Confluence Model of sexual aggression to predict men's conflict with women: A 10-year follow-up study. *Journal of Personality and Social Psychology*, *69(2)*, 353–369. https://psycnet.apa.org/buy/1995-44088-001

Malamuth, N. M., Sockloskie, R. J., Koss, M. P., & Tanaka, J. S. (1991). Characteristics of aggressors against women: Testing a model using a national sample of college students. *Journal of Consulting and Clinical Psychology*, *59*(5), 670–681. https://doi.org/10.1037/0022-006X.59.5.670

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*(3), 257–284. https://doi.org/10.1037/a0032773

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*(1), 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

Matsunaga, M. (2008). Item parceling in structural equation modeling: A Primer. *Communication Methods and Measures*, *2*(4), 260–293. https://doi.org/10.1080/19312450802458935

McDonald, R. P. (1999). *Test theory: A unified treatment.* Lawrence Erlbaum Associates.

Milaniak, I., & Widom, C. S. (2015). Does child abuse and neglect increase risk for perpetration of violence inside and outside the home? *Psychology of Violence*, *5*(3), 246–255. https://doi.org/10.1037/a0037956

Mouilso, E., & Calhoun, K. (2013). The role of rape myth acceptance and psychopathy in sexual assault perpetration. *Journal of Aggression Maltreatment & Trauma*, *22*(2), Article 2. https://doi.org/10.1080/10926771.2013.743937

Najdowski, C., & Ullman, S. E. (2009). PTSD symptoms and self-rated recovery among adult sexual assault survivors: The effects of traumatic life events and psychosocial variables. *Psychology of Women Quarterly, 33*(1), 43-53. https://doi.org/10.1111/j.1471-6402.2008.01473.x

Nelson, P. A. (1979). *Personality, sexual functions, and sexual behavior: An experiment in methodology* (Vol. 39, p. 6134). ProQuest Information & Learning.

Nering, M., & Ostini, R. (2011). *Handbook of Polytomous Item Response Theory Models*. Taylor & Francis.

Nguyen, D., & Parkhill, M. R. (2014). Integrating attachment and depression in the Confluence Model of sexual assault perpetration. *Violence Against Women*, *20*(8), 994–1011. https://doi.org/10.1177/1077801214546233

Ogle, R. L., Noel, N. E., & Maisto, S. A. (2009). Assessing acceptance of violence toward women: A factor analysis of Burt's acceptance of interpersonal violence scale. *Violence Against Women*, *15*(7), 799–809. https://doi.org/10.1177/1077801209334444

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied psychological measurement*, *24*(1), 50-64. https://doi-org/10.1177/01466216000241003

Owens, B. C., Hall, M. E. L., & Anderson, T. L. (2021). The relationship between purity culture and rape myth acceptance. *Journal of Psychology and Theology*, *49*(4), 405–418. https://doi.org/10.1177/0091647120974992

Parkhill, M. R., & Abbey, A. (2008). Does alcohol contribute to the confluence model of sexual assault perpetration? *Journal of Social and Clinical Psychology*, *27*(6), 529–554. https://doi.org/10.1521/jscp.2008.27.6.529

Penfield, R. D. (2014). An NCME instructional module on polytomous item response theory models. *Educational Measurement: Issues and Practice*, *33*(1), 36–48. https://doi.org/10.1111/emip.12023

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Peterson, C., DeGue, S., Florence, C., & Lokey, C. N. (2017). Lifetime economic burden of rape among U.S. adults. *American Journal of Preventive Medicine*, *52*(6), 691–701. https://doi.org/10.1016/j.amepre.2016.11.014

Rabaud, E., Bègue, L., & Bushman, B. J. (2025). Alcohol consumption and sexual aggression perpetration: A meta-analytic review. *Trauma, Violence, & Abuse*, 15248380251366266. https://doi.org/10.1177/15248380251366266

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Ray, T. N., & Parkhill, M. R. (2021). Components of hostile masculinity and their associations with male-perpetrated sexual aggression toward women: A systematic review. *Trauma, Violence, & Abuse*, 152483802110302. https://doi.org/10.1177/15248380211030224

Raykov, T. (1998). A method for obtaining standard errors and confidence intervals of composite reliability for congeneric items. *Applied Psychological Measurement*, *22*(4), 369–374. https://doi.org/10.1177/014662169802200406

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2013). Applying unidimensional item response theory models to psychological data. In *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 101–119). American Psychological Association. https://doi.org/10.1037/14047-006

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, *27*(2), 133–144. https://doi.org/10.1111/j.1745-3984.1990.tb00738.x

Rezapour, M., Cuccolo, K., Veenstra, C., & Ferraro, F. R. (2021). An item response theory to analyze the psychological impacts of rail-transport delay. *Sustainability*, *13*(12), 6935. https://doi.org/10.3390/su13126935

Rhemtulla, M. (2016). Population performance of SEM parceling strategies under measurement and structural model misspecification. *Psychological Methods*, *21*(3), 348–368. https://doi.org/10.1037/met0000072

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*(4, Pt. 2), 100–100.

Santaularia, J., Johnson, M., Hart, L., Haskett, L., Welsh, E., & Faseru, B. (2014). Relationships between sexual violence and chronic disease: A cross-sectional study. *BMC Public Health*, *14*(1), 1286. https://doi.org/10.1186/1471-2458-14-1286

Saqib, S., & Davidson, M. (2023). The Confluence Model of sexual aggression: The role of pornography as a secondary risk factor. *Sexual Offending: Theory, Research, and Prevention*, *18*, 1–24. https://doi.org/10.5964/sotrap.13005

Sass, D. A., & Smith, P. L. (2006). The effects of parceling unidimensional scales on structural parameter estimates in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(4), 566–586. https://doi.org/10.1207/s15328007sem1304_4

Schermelleh-engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 23–74.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321. https://doi.org/10.1177/0734282911406653

Schwartz, I. L. (1991). Sexual violence against women: Prevalence, consequences, societal factors, and prevention. *American Journal of Preventive Medicine*, *7*(6), 363–373. https://doi.org/10.1016/S0749-3797(18)30873-0

Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., Gundy, C., Koller, M., Petersen, M. A., & Sprangers, M. A. G. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, *62*(3), 288–295. https://doi.org/10.1016/j.jclinepi.2008.06.003

Seibert, L. A., Miller, J. D., Few, L. R., Zeichner, A., & Lynam, D. R. (2011). An examination of the structure of self-report psychopathy measures and their relations with general traits and externalizing behaviors. *Personality Disorders: Theory, Research, and Treatment*, *2*(3), 193–208. https://doi.org/10.1037/a0019232

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428–1441. https://doi.org/http://dx.doi.org.huaryu.kl.oakland.edu/10.1037/pas0000623

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, *25*(3). https://doi.org/10.1214/10-STS330

Simpson, J. A., & Gangestad, S. W. (1991). Individual differences in sociosexuality: Evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology*, *60*(6), 870–883. http://search.proquest.com/docview/80725806?rfr_id=info%3Axri%2Fsid%3Aprimo

Smith, L., George, W. H., & Neilson, E. C. (2024). The Confluence Model of sexual aggression in the context of acute intoxication and state emotion regulation. *Violence and Victims*, *39*(5), 571–587. https://doi.org/10.1891/VV-2022-0131

Smith, R. M., Parrott, D. J., Swartout, K. M., & Tharp, A. T. (2015). Deconstructing hegemonic masculinity: The roles of antifemininity, subordination to women, and sexual dominance in men's perpetration of sexual aggression. *Psychology of Men & Masculinity*, *16*(2), 160–169. https://doi.org/10.1037/a0035956

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE*, *8*(7), e68967. https://doi.org/10.1371/journal.pone.0068967

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment*, *91*(1), 62–71. https://doi.org/10.1080/00223890802484381

Sterba, J. P. (2019). *Is a Good God Logically Possible?* Springer.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Sugar, N. F., Fine, D. N., & Eckert, L. O. (2004). Physical injury after sexual assault: Findings of a large case series. *American Journal of Obstetrics and Gynecology*, *190*(1), 71–76. https://doi.org/10.1016/S0002-9378(03)00912-8

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116–131. https://doi.org/10.1509/jmkr.45.1.116

Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research*, *16*(S1), 33–42. https://doi.org/10.1007/s11136-007-9184-6

Tharp, A. T., DeGue, S., Valle, L. A., Brookmeyer, K. A., Massetti, G. M., & Matjasko, J. L. (2013). A systematic qualitative review of risk and protective factors for sexual violence perpetration. *Trauma, Violence, & Abuse*, *14*(2), 133–167. https://doi.org/10.1177/1524838012470031

Ullman, J. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment*, *87*, 35–50. https://doi.org/10.1207/s15327752jpa8701_03

Ullman, S., Relyea, M., Peter-Hagene, L., & Vasquez, A. (2013). Trauma histories, substance use coping, PTSD, and problem substance use among sexual assault victims. *ADDICTIVE BEHAVIORS*, *38*(6), Article 6. https://doi.org/10.1016/j.addbeh.2013.01.027

Ursoniu, S., Serban, C. L., Giurgi-Oncu, C., Rivis, I. A., Bucur, A., Bredicean, A.-C., & Papava, I. (2021). Validation of the Romanian version of the Toronto Empathy Questionnaire (TEQ) among undergraduate medical students. *International Journal of Environmental Research and Public Health*, *18*(24), 12871. https://doi.org/10.3390/ijerph182412871

Vandello, J. A., & Bosson, J. K. (2013). Hard won and easily lost: A review and synthesis of theory and research on precarious manhood. *Psychology of Men & Masculinity*, *14*(2), 101–113. https://doi.org/10.1037/a0029826

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*(3), 320–334. https://doi.org/10.1037/a0032121

Wheeler, J. G., George, W. H., & Dahl, B. J. (2002). Sexually aggressive college males: Empathy as a moderator in the "Confluence Model" of sexual aggression. *Personality and Individual Differences*, *33*(5), 759–775. https://doi.org/10.1016/S0191-8869(01)00190-8

Whitmire, L. E., Harlow, L. L., Quina, K., & Morokoff, P. J. (1999). *Childhood trauma and HIV: Women at risk.* Philadelphia, PA: Brunner/Mazel.

Williams, K. M., Paulhus, D. L., & Hare, R. D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of Personality Assessment*, *88*(2), 205–219. https://doi.org/10.1080/00223890701268074

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/10.3354/cr030079

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52–69. https://doi.org/10.20982/tqmp.08.1.p052

Woods, C. M. (2008). Likelihood-ratio DIF testing: Effects of nonnormality. *Applied Psychological Measurement*, *32*(7), 511–526. https://doi.org/10.1177/0146621607310402

Wright, P. J., Paul, B., & Herbenick, D. (2021). Pornography, impersonal sex, and sexual aggression: A test of the confluence model in a national probability sample of men in the U.S. *Aggressive Behavior*, *47*(5), 593–602. https://doi.org/10.1002/ab.21978

Yucel, E., Angelone, D. J., & Jones, M. C. (2023). Reassessing the Confluence Model of men's risk for sexual aggression. *Journal of Interpersonal Violence*, *38*(7–8), 6062–6084. https://doi.org/10.1177/08862605221127376

Zein, R. A., & Akhtar, H. (2025). Getting started with the graded response model: An introduction and tutorial in R. *International Journal of Psychology*, *60*(1), e13265. https://doi.org/10.1002/ijop.13265

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/http://dx.doi.org.huaryu.kl.oakland.edu/10.1007/s11336-003-0974-7

Zurbriggen, E. L. (2000). Social motives and cognitive power–sex associations: Predictors of aggressive sexual behavior. *Journal of Personality and Social Psychology*, *78*(3), 559–581. https://doi.org/10.1037/0022-3514.78.3.559