

# Gender Dynamics in Rape Myth Acceptance: A Psychometric Evaluation of the IRMA-S “She Asked For It” Subscale in Spanish

Carmen M. Leon<sup>1</sup>, Eva Aizpurua<sup>2</sup>, Tatiana Quiñonez-Toral<sup>1</sup>

[1] School of Law, University of Castilla-La Mancha, Albacete, Spain. [2] Centre for Social Survey Transformation, National Centre for Social Research, London, United Kingdom.

---

Measurement Instruments for the Social Sciences, 2026, Vol. 8, Article e19377, <https://doi.org/10.5964/miss.19377>

---

**Received:** 2025-08-18 • **Accepted:** 2026-04-16 • **Published (VoR):** 2026-06-11

---

**Handling Editor:** Piotr Koc, GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany

---

**Corresponding Author:** Carmen M. Leon, 1 Plaza de la Universidad, 02008 Albacete, Spain. E-mail: [Carmen.Leon@uclm.es](mailto:Carmen.Leon@uclm.es)

---

**Supplementary Materials:** Materials [see [Index of Supplementary Materials](#)]



## Abstract

Although men consistently endorse rape myths at higher rates than women, valid gender comparisons require evidence that instruments function equivalently across groups. This study evaluates the psychometric properties and gender-based measurement invariance of the “She Asked For It” subscale from the *Illinois Rape Myth Acceptance Scale-Subtle* (IRMA-S) in a Spanish community sample ( $N = 1,003$ ; 50.7% men). Analyses included Multi-Group Confirmatory Factor Analysis and Item Response Theory-based approaches, specifically Differential Item Functioning (DIF) and Differential Response Functioning (DRF). At the subscale level, analyses supported metric invariance and satisfactory internal consistency. However, five of six items exhibited significant DIF, predominantly uniform in nature. DRF analyses indicated that the cumulative impact of item-level DIF on expected total scores was modest, particularly for respondents with moderate levels of rape myth acceptance. These findings highlight the value of multi-method approaches to measurement invariance and underscore the need for cautious interpretation of raw gender differences in rape myth acceptance.

## Keywords

comparability, gender, online survey, psychometric evaluation, rape myth acceptance, Spanish



Understanding attitudes toward sexual violence is crucial for effective prevention efforts. Rape Myth Acceptance (RMA) refers to the endorsement of stereotypical or misleading beliefs about rape, victims, or perpetrators that justify sexual aggression, minimize its severity, or shift blame from the perpetrator to the victim (Burt, 1980). Because attitudes toward sexual violence are shaped by evolving cultural and social norms, it is essential to maintain valid and up-to-date instruments that allow for reliable comparisons across populations and over time.

Building upon this need, the current study examines the Spanish adaptation of the “She Asked For It” subscale from the most recent refinement of the *Illinois Rape Myth Acceptance Scale*—the IRMA-Subtle (IRMA-S; Thelan & Meadows, 2021). This version represents the latest step in the iterative refinement of the IRMA framework, designed to enhance inclusivity, cross-cultural generalizability, and sensitivity to contemporary expressions of RMA. Specifically, we assess the psychometric properties of the adapted subscale by examining subscale-level measurement invariance, item-level Differential Item Functioning (DIF), and the scale-level impact of DIF across gender to evaluate its cross-group comparability.

## Development of the Illinois Rape Myth Acceptance Scale

The development of rape myth acceptance measures has been central to understanding how individuals and societies conceptualize and rationalize sexual violence. Foundational work began with Burt’s (1980) Rape Myth Acceptance Scale (RMAS), which assessed culturally embedded attitudes about sexual assault. The scale primarily targeted victim-blaming attitudes and reflected the prevailing social norms of the late 1970s, including items such as “Many women secretly desire to be raped.”

Recognizing the limitations of such explicit phrasing, Payne et al. (1999) developed the Illinois Rape Myth Acceptance (IRMA) Scale and its short form (IRMA-SF). To capture the multidimensional nature of RMA, this instrument introduced seven theoretically grounded subscales: (1) “She asked for it,” (2) “It wasn’t really rape,” (3) “He didn’t mean to,” (4) “She wanted it,” (5) “She lied,” (6) “Rape is a trivial event,” and (7) “Rape is a deviant event.” This structure allowed for a more nuanced assessment of the attitudinal dimensions underlying RMA.

Subsequent revisions followed an iterative process aimed at enhancing conceptual clarity, psychometric robustness, and cultural relevance. McMahon and Farmer (2011) revised the IRMA for college populations (IRMA-2011), reducing the original seven subscales to four (“She asked for it,” “It wasn’t really rape,” “He didn’t mean to,” and “She lied”<sup>1</sup>) and rewording items to capture more subtle and socially acceptable expressions of RMA, particularly those rooted in contemporary victim-blaming rhetoric. For example,

---

1) “*She asked for it*” reflects the belief that the victim’s behaviors invited the sexual assault. “*It wasn’t really rape*” denies that an assault occurred due to either blaming the victim or excusing the perpetrator. “*He didn’t mean to*”

the original item “When women go around wearing low-cut tops or short skirts, they’re just asking for trouble” was revised to “When girls go to parties wearing slutty clothes, they are asking for trouble.”

Over time, research has documented a decline in the explicit endorsement of rape myths (Kleinsmith et al., 2026; Leon et al., 2025), yet more subtle and implicit myths continue to shape public attitudes. Building on this evolution, Thelan and Meadows (2021) developed the IRMA-Subtle (IRMA-S), representing the latest refinement of the instrument. The IRMA-S retains all 22 items introduced in the IRMA-2011 while updating wording to enhance inclusivity, cross-cultural generalizability, and applicability beyond college populations. For instance, the item “When girls go to parties wearing slutty clothes, they are asking for trouble” was reformulated as “When women go out wearing slutty clothes, they are asking for sexual advances from men.”

## Gender-Based Variability in Rape Myth Acceptance Across Cultural Contexts

Studies conducted across diverse cultural contexts generally show that men endorse rape myths at higher rates than women, revealing robust gender differences in RMA (Fansher & Zedaker, 2022; Kazmi et al., 2024; Martini et al., 2022). However, determining whether these differences reflect genuine attitudinal variation or measurement artifacts requires careful examination of the comparability of the instruments used.

Measurement invariance refers to the extent to which a scale measures the same underlying construct across groups or contexts (Dong & Dumas, 2020). Establishing measurement invariance provides evidence that observed gender differences reflect true disparities in beliefs rather than systematic differences in interpretation or response patterns. Without such evidence, cross-group comparisons may be misleading and risk conflating substantive variation with measurement artifacts.

Empirical research on the IRMA and its adaptations has yielded mixed findings regarding gender-based invariance. In Russia, Balezina and Zakharova (2023) reported that the IRMA-SF demonstrated invariance across men and women. Similarly, the Hungarian adaptation of the IRMA-2011 supported gender invariance, suggesting comparable functioning across groups (Nyúl & Kende, 2023). Genc et al. (2026), using a Serbian version of the IRMA-S, found that men scored significantly higher than women, yet their analyses supported measurement invariance, indicating comparable scale functioning. In contrast, Fakunmoju et al. (2019), employing the English IRMA-2011 in Nigeria, found that although men scored higher across subscales, only 16 of the 22 items met invariance criteria across gender.

---

reflects the idea that the perpetrator did not intend to rape. “*She lied*” indicates the belief that the victim fabricated the rape (McMahon & Farmer, 2011).

Despite consistent evidence of gender differences in raw scores, systematic testing of measurement invariance—particularly at the item level—remains relatively limited in both English-language and translated versions of the IRMA. Establishing measurement invariance is therefore a prerequisite for interpreting observed gender differences as substantive attitudinal variation rather than potential measurement artifacts, highlighting the importance of formally testing invariance when adapting RMA measures across contexts.

## The Current Study

To date, no version of the IRMA has been formally validated in Spain. Although a Spanish translation of the IRMA-2011 was validated in Argentina (González-Caino et al., 2022), cultural and linguistic differences limit its direct applicability to the Spanish context. In Spain, selected IRMA-2011 items have appeared in official surveys (e.g., Delegación del Gobierno Contra la Violencia de Género, 2017); however, they were administered alongside items from other instruments, precluding a comprehensive psychometric evaluation of the scale as a coherent measure.

Additionally, while the *Acceptance of Modern Myths about Sexual Aggression* (AMM-SA; Megías et al., 2011) is widely used in Spain, its unidimensional structure and length (30 items) make it less suitable for multidimensional analyses or survey contexts where brevity is required. By contrast, the IRMA-S provides a concise and theoretically structured framework that distinguishes specific domains of RMA. Notably, the “She Asked For It” subscale captures victim-blaming beliefs linked to perceived provocation, which remain particularly salient in Spain—especially following the *La Manada*<sup>2</sup> case and subsequent national debate surrounding consent and victim responsibility.

Spain therefore represents a highly relevant sociocultural setting in which to examine the functioning of this subscale, given increased institutional attention to sexual violence and the implementation of the 2022 “Solo Sí es Sí”<sup>3</sup> law.<sup>4</sup> Although these developments have shaped public discourse and may influence the social acceptability of certain rape

---

2) *La Manada* refers to a 2016 gang rape case in Pamplona, Spain, in which five men sexually assaulted an 18-year-old woman during the San Fermín festival. The initial 2018 court ruling convicted the perpetrators of sexual abuse rather than rape, triggering widespread public protests and national debate over Spain’s legal definitions of sexual violence and consent.

3) In response to public outcry and subsequent legal developments, the Spanish government enacted the “Solo Sí es Sí” (“Only Yes Means Yes”) law, which centers affirmative consent in sexual offense legislation and aims to eliminate distinctions between sexual abuse and assault.

4) The focus on this specific subscale is also consistent with the broader objectives of the survey from which the data were drawn. The questionnaire was part of a larger study examining public perceptions of sexual violence, including attitudes related to the attribution of responsibility to victims across different scenarios. Within this framework, the “She Asked For It” subscale was particularly relevant because it directly captures victim-blaming beliefs linking women’s behavior or appearance to sexual victimization, making it especially suitable for examining the attitudinal dimensions targeted in the research.

myths, the primary contribution of the current study is methodological: addressing the persistent lack of systematic testing of gender-based measurement invariance in instruments assessing RMA.

Accordingly, this study evaluates the psychometric properties of the Spanish version of the “She Asked For It” subscale, with particular focus on subscale- and item-level measurement invariance across gender. To this end, Multi-Group Confirmatory Factor Analysis (MG-CFA) and Item Response Theory (IRT)-based approaches, specifically Differential Item Functioning (DIF) and Differential Response Functioning (DRF), are employed. This multi-method approach allows for a rigorous evaluation of whether the subscale functions comparably across genders (Zumbo, 2003).

## Method

### Participants and Procedure

Data were collected online between January 19 and February 14, 2024, using Verian’s opt-in panel in Spain. Gender and age quotas were applied to approximate the demographic distribution of the population of Castilla-La Mancha (see Supplementary Material for comparison with regional census data on key variables). The survey completion rate was 36.1%. Respondents received points redeemable for gifts as compensation for their time.

The final sample comprised 1,003 residents of Castilla-La Mancha (50.7% men), with ages ranging from 16 to 84 years ( $M = 45.5$ ,  $SD = 14.3$ ). The survey was administered in Spanish and included up to 57 items assessing public attitudes toward sexuality, including measures of RMA and ambivalent sexism, among other constructs. The average completion time was approximately 12 minutes ( $M = 12.2$ ,  $mdn = 10.9$ ,  $SD = 5.8$ ). Ethical approval for the study was granted by the Social Research Ethics Committee at the University of Castilla-La Mancha (Protocol Number: CEIS-2024-23586).

### Measures

#### Illinois Rape Myth Acceptance Scale – Subtle (IRMA-S)

Rape myth acceptance was assessed using the “She Asked For It” subscale of the IRMA-S, developed and validated by Thelan and Meadows (2021) in the US. The subscale comprises six items reflecting the belief that certain behaviors or appearances on the part of victims invite sexual assault. Responses were recorded on a 5-point Likert-type scale ranging from 1 (*totally disagree*) to 5 (*totally agree*). The subscale showed satisfactory internal consistency ( $\alpha = .80$ ).

Items from the “She Asked For It” subscale were translated following the TRAPD method (Translation, Review, Adjudication, Pretesting, and Documentation). Two independent translators with expertise in sexual violence research produced initial transla-

tions, which were reviewed by a third expert who reconciled discrepancies and generated a consolidated version. The reconciled version was subsequently pretested, along with other questionnaire items, by a panel of four experts (three specialists in sexual violence and one psychometrician).

The expert panel evaluated semantic clarity and interpretability using a structured review template (see Supplementary Material). Based on their feedback, minor linguistic refinements were introduced to reduce potential ambiguity. The research team concluded that the revised wording achieved adequate conceptual equivalence and was appropriate for self-administration. All stages of the translation and adaptation process were documented to ensure methodological transparency. The final wording of the items, in both the original English and Spanish-adapted versions, is available in the Supplementary Material.

## Analytic Strategy

Statistical analyses were conducted in *R* (version 4.2.3). Descriptive statistics were computed for all items in the “She Asked For It” subscale. In line with current recommendations in psychometric research, measurement invariance was evaluated at both subscale and item levels. Although MG-CFA provides a well-established framework for assessing invariance at the (sub)scale level, it may be insensitive to non-invariance affecting specific items (D’Urso et al., 2022). As a result, item-level non-invariance may exist even when (sub)scale-level indices suggest acceptable model fit. IRT-based approaches address this limitation by enabling the detection of DIF, that is, whether individual items operate differently across groups after controlling for the latent trait (Tay et al., 2014). For this reason, combining MG-CFA and IRT-based approaches provides a more comprehensive and sensitive assessment of measurement equivalence.

Subscale-level measurement invariance across gender was examined using MG-CFA implemented in the *lavaan* and *semTools* packages. Given the ordinal nature of the response scale (1–5), the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator was employed. WLSMV provides appropriate parameter estimates, standard errors, and scaled chi-square statistics for ordinal indicators.

Following recommendations for ordered-categorical indicators (Wu & Estabrook, 2016; Svetina et al., 2019), measurement invariance was tested sequentially across three levels: configural invariance (equal factor structure across groups with freely estimated thresholds and loadings), threshold invariance (thresholds constrained equal across groups), and metric invariance (thresholds and factor loadings constrained equal). Model fit was evaluated using the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Nested model comparisons were conducted using scaled chi-square difference tests appropriate for WLSMV estimation, supplemented by changes in CFI and RMSEA.

Internal consistency was assessed using Cronbach's alpha, ordinal alpha, and McDonald's omega coefficients, calculated with the *semTools* package. Ordinal reliability coefficients were computed to account for the ordered-categorical nature of the data (Zumbo et al., 2007).

Item-level performance was assessed using the *mirt* package. A multi-group Graded Response Model (MG-GRM) was estimated to model the relationship between the latent trait (RMA) and the probability of endorsing each response category across gender groups. MG-GRMs yield item discrimination parameters ( $a$ ), which indicate how strongly each item relates to the latent construct, and category difficulty parameters ( $\beta$ ), which represent locations along the latent trait continuum at which respondents have a 50% probability of endorsing a given response category or higher (i.e., cumulative category probabilities).

Item-level measurement invariance was further evaluated using Differential Item Functioning (DIF) analyses within this multi-group IRT framework. For each item, a sequential series of likelihood ratio chi-square tests was conducted. First, an omnibus DIF test compared a constrained baseline model with a model in which both discrimination ( $a$ ) and category difficulty ( $\beta$ ) parameters were freely estimated across gender. Second, non-uniform DIF was evaluated by testing equality constraints on discrimination parameters. Third, uniform DIF was assessed by testing equality constraints on category difficulty parameters while constraining discrimination parameters to equality across groups. This sequential procedure allowed differentiation between non-uniform DIF attributable to differences in item discrimination and uniform DIF attributable to differences in category difficulty parameters. This approach complements MG-CFA by providing item-level evidence of measurement non-equivalence. Whereas MG-CFA evaluates invariance at the subscale level, DIF analyses identify specific items that may operate differently across groups, even when overall subscale-level invariance is supported.

To assess the cumulative impact of item-level DIF on total scores, DRF analyses were conducted using the *mirt* package (Chalmers, 2018). DRF examines whether expected total scores differ between groups at equivalent levels of the latent trait across the full continuum. To incorporate sampling uncertainty into the presentation of results, 95% confidence intervals were obtained using 500 bootstrap draws for both item-level and scale-level DRF estimates.

## Results

### Rape Myth Acceptance

Overall, respondents exhibited low endorsement of rape myth statements. Table 1 presents the percentage distribution across response categories for each of the six items. Most respondents selected "totally disagree" or "quite disagree," with pronounced floor

effects for Items 1 (“If a woman is raped while she is drunk, she is responsible for what happened”) and 5 (“When women are raped, it is often because the way they said ‘no’ was unclear”).

**Table 1**

*Item-Level Response Distribution for the “She Asked For It” Subscale Across the Full Sample (N = 1,003) and by Gender*

Item	Totally disagree % (n)	Quite disagree % (n)	Neither agree nor disagree % (n)	Quite agree % (n)	Totally agree % (n)
<b>If a woman is raped while she is drunk, she is responsible for what happened / Si violan a una mujer estando borracha, ella es responsable de lo sucedido</b>					
Full sample	87.1% (874)	7.4% (74)	4.1% (41)	1.0% (10)	0.4% (4)
Men	85.9% (437)	9.0% (46)	3.5% (18)	1.4% (7)	0.2% (1)
Women	88.5% (437)	5.7% (28)	4.7% (23)	0.6% (3)	0.6% (3)
<b>When women go out wearing slutty clothes, they are asking for sexual advances from men / Cuando las mujeres se visten con ropa provocativa, están buscando que los hombres se les insinúen</b>					
Full sample	60.9% (611)	16.5% (165)	14.1% (141)	6.9% (69)	1.7% (17)
Men	48.1% (245)	22.4% (114)	18.7% (95)	8.6% (44)	2.2% (11)
Women	74.1% (366)	10.3% (51)	9.3% (46)	5.1% (25)	1.2% (6)
<b>If a woman goes home with a man after a party, it is her own fault if she has sex and does not want to / Si una mujer se va a casa con un hombre después de una fiesta, es su culpa si acaba teniendo relaciones sexuales que no quería</b>					
Full sample	68.4% (686)	14.3% (143)	11.7% (117)	4.1% (41)	1.6% (16)
Men	61.5% (313)	17.3% (88)	14.3% (73)	5.3% (27)	1.6% (8)
Women	75.5% (373)	11.1% (55)	8.9% (44)	2.8% (14)	1.6% (8)
<b>If a woman sleeps around, eventually something bad is going to happen to her / Si una mujer es promiscua, terminará pasándole algo malo</b>					
Full sample	49.2% (493)	21.2% (213)	21.6% (217)	5.7% (57)	2.3% (23)
Men	41.7% (212)	23.0% (117)	25.5% (130)	7.3% (37)	2.6% (13)
Women	56.9% (281)	19.4% (96)	17.6% (87)	4.1% (20)	2.0% (10)
<b>When women are raped, it is often because the way they said ‘no’ was unclear / A menudo, cuando las mujeres son violadas, es porque la forma en que dijeron ‘no’ fue poco clara</b>					
Full sample	81.3% (815)	11.0% (110)	6.3% (63)	0.8% (8)	0.7% (7)
Men	75.8% (386)	14.2% (72)	8.1% (41)	1.2% (6)	0.8% (4)
Women	86.8% (429)	7.7% (38)	4.5% (22)	0.4% (2)	0.6% (3)
<b>If a woman starts making out, she should not be surprised if a man assumes she wants to have sex / Si una mujer empieza a besar a un hombre, no debería sorprenderse si él asume que ella quiere tener sexo</b>					
Full sample	47.0% (471)	18.7% (188)	18.3% (184)	11.8% (118)	4.2% (42)
Men	33.0% (168)	21.2% (108)	24.2% (123)	16.7% (85)	4.9% (25)
Women	61.3% (303)	16.2% (80)	12.4% (61)	6.7% (33)	3.4% (17)

*Note.* Significant differences between men and women were found for all items comprising the subscale. Results available upon request. Full Sample (N = 1,003); Men (n = 509); Women (n = 494).

Gender differences in response distributions appeared minimal for Item 1. In contrast, for the remaining items—particularly Items 2 (“When women go out wearing slutty clothes, they are asking for sexual advances from men”) and 6 (“If a woman starts making out, she should not be surprised if a man assumes she wants to have sex”)—women showed higher rates of disagreement and lower endorsement of agreement categories compared to men (see Table 1).

## Subscale-Level Measurement Invariance Across Gender

The single-factor model demonstrated excellent fit for women (scaled  $\chi^2(8) = 8.93$ ,  $p = .348$ ; CFI = 1.000; TLI = .999; RMSEA = .015; SRMR = .017), whereas model fit was comparatively weaker for men (scaled  $\chi^2(8) = 41.20$ ,  $p < .001$ ; CFI = 0.986; TLI = 0.974; RMSEA = .090; SRMR = .046). Although the chi-square test was significant for men, incremental fit indices and SRMR suggested acceptable though comparatively weaker model fit.

In the male subsample, a high modification index (MI = 28.59) was observed for the residual correlation between Items 1 (“If a woman is raped while she is drunk, she is responsible for what happened”) and 5 (“When women are raped, it is often because the way they said ‘no’ was unclear”) (Expected Parameter Change = -0.26; standardized EPC = -0.742). Because both items belong to the same subscale and address closely related victim-blaming content, this residual correlation—substantially stronger for men ( $r = .474$ ) than women ( $r = .268$ )—likely reflects local item dependence not fully captured by the latent factor.

To account for this localized dependency while preserving comparability across groups, the model was re-specified to allow the residuals of these two items to correlate in both gender groups. Formal tests of measurement invariance were conducted following Wu and Estabrook’s (2016) recommendations for categorical indicators. Table 2 presents fit indices and model comparisons across the invariance sequence.

**Table 2**

*Subscale-Level Measurement Invariance Across Gender*

Model	$\chi^2$ (scaled)	$df$	$p$	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$	$\Delta df$	$p(\Delta\chi^2)$	$\Delta CFI$	$\Delta RMSEA$
Configural	55.135	16	< .001	.991	.983	.070	.031	—	—	—	—	—
Threshold	66.111	28	< .001	.991	.990	.052	.031	9.647	12	.647	.000	-.018
Metric	57.191	33	.006	.994	.995	.038	.032	2.801	5	.731	.003	-.014

*Note.* All fit indices based on robust (scaled) statistics using WLSMV estimator. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

The configural model demonstrated acceptable fit (scaled  $\chi^2(16) = 55.14, p < .001$ ; CFI = .991; TLI = .983; RMSEA = .070; SRMR = .031). Although the chi-square test was significant, incremental fit indices and SRMR indicated satisfactory model fit. Threshold invariance, which constrains item thresholds to equality across gender, showed no significant decrement in fit relative to the configural model ( $\Delta\chi^2(12) = 9.65, p = .647$ ). Fit indices remained stable (CFI = .991; TLI = .990; RMSEA = .052; SRMR = .031), and changes in CFI ( $\Delta\text{CFI} = .000$ ) and RMSEA ( $\Delta\text{RMSEA} = -.018$ ) were well within commonly recommended criteria for supporting invariance.

Metric invariance, which constrains both thresholds and factor loadings to equality, similarly showed no significant worsening of fit relative to the threshold model ( $\Delta\chi^2(5) = 2.80, p = .731$ ). Model fit did not deteriorate under metric constraints, providing support for metric invariance at the subscale level across gender. Changes in fit indices across models were minimal and remained within recommended thresholds (see Table 2). Taken together, these findings support metric invariance at the subscale level, indicating that the latent construct operates comparably across genders within the MG-CFA framework.

Standardized factor loadings from the metric invariance model ranged from .730 to .871, indicating strong associations between all items and the latent construct. Reliability estimates for the re-specified model were satisfactory across the full sample and within each gender subgroup. Cronbach's alpha indicated acceptable internal consistency ( $\alpha = .80$  overall;  $\alpha = .81$  for women;  $\alpha = .78$  for men). As expected for ordinal data, ordinal alpha coefficients were higher ( $\alpha_{ord} = .91$  for women;  $\alpha_{ord} = .87$  for men), and McDonald's omega values also indicated robust reliability ( $\omega = .83$  overall;  $\omega = .84$  for women;  $\omega = .80$  for men).

## Item-Level Measurement Invariance Across Gender

Item-level DIF was assessed within a multi-group IRT framework. Although subscale-level MG-CFA supported metric invariance, DIF analyses revealed non-invariance. Omnibus DIF tests indicated significant DIF in five of the six items (see Table 3), with only Item 1 demonstrating invariance across gender ( $\chi^2(5) = 6.262, p = .282$ ). Tests of non-uniform DIF were non-significant for all items except Item 2, for which a statistically detectable difference in discrimination was observed ( $\chi^2(1) = 5.03, p = .025$ ). In contrast, tests of uniform DIF were highly significant for Items 2, 3, 4, and 6 (all  $p < .001$ ) and marginally significant for Item 5 ( $\chi^2(4) = 9.49, p = .050$ ).

These results indicate that item-level non-invariance was driven primarily by differences in category difficulty rather than item discrimination. Across most items, men and women with equivalent levels of latent RMA differed systematically in their probability of endorsing higher response categories, while item discrimination was largely comparable across groups.

**Table 3***Differential Item Functioning Tests Across Gender*

Item	Omnibus DIF $\chi^2(df)$	Non-uniform DIF $\chi^2(df)$	Uniform DIF $\chi^2(df)$
1	6.26 (5)	0.39 (1)	6.00 (4)
2	75.30***	5.03*	74.46***
3	23.17***	0.81	20.25***
4	23.42***	1.71	23.41***
5	19.47**	0.02	9.49*
6	89.23***	0.89	88.94***

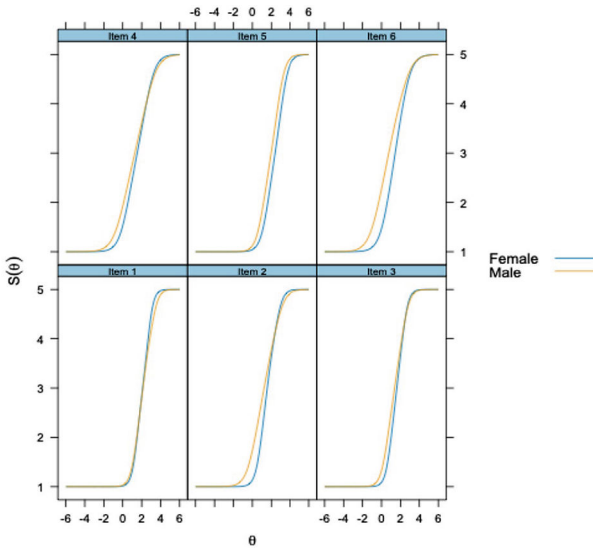
*Note.* Omnibus DIF tests jointly constrained discrimination and category difficulty parameters across gender. Non-uniform DIF tests constrained discrimination parameters only. Uniform DIF tests constrained category difficulty parameters while allowing discrimination parameters to vary.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

Figures 1a and 1b present expected item score functions and the corresponding expected score differences across the latent trait continuum. At equivalent trait levels, men generally exhibit higher expected item scores than women, particularly for Items 2 and 6. This pattern reflects a leftward shift in the expected score functions for men, indicating that lower levels of the latent trait are required for men to endorse higher response categories relative to women. For example, at  $\theta = 0$  (the mean of the latent trait), Item 6 yields an expected score of approximately 3 for men compared to approximately 2 for women, representing roughly a one-point difference on the 5-point response scale. In contrast, Item 1 shows overlapping expected score functions across groups, consistent with the absence of DIF.

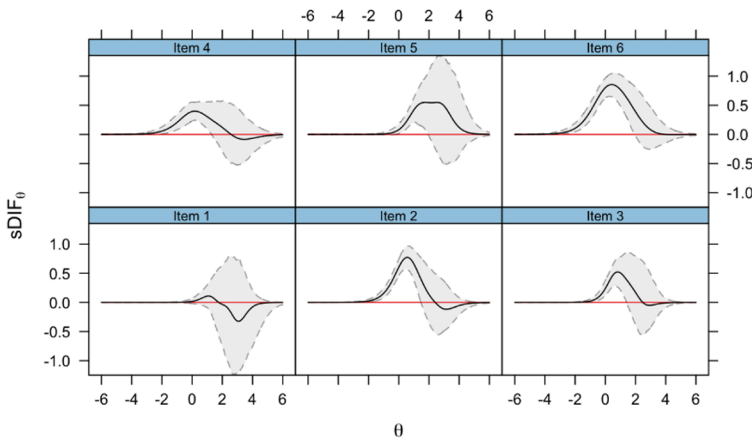
**Figure 1a**

*Expected Item Scores by Gender from the Multi-Group Graded Response Model*



**Figure 1b**

*Differences in Expected Item Scores Between Men and Women Across the Latent Trait Continuum*



*Note.* The solid black line represents the point estimate of the male–female difference at each level of the latent trait ( $\theta$ ); the shaded grey band represents the 95% confidence interval derived from 500 bootstrap draws. The red horizontal reference line indicates no difference. Positive values indicate higher expected item scores for men than women at equivalent levels of latent rape myth acceptance.

Examination of item parameter estimates (see Table 4) further clarified the nature of the detected DIF. Discrimination parameters ( $a$ ) were generally higher for women, indicating steeper item response functions and greater differentiation of the latent trait among female respondents. For example, the discrimination parameter for Item 2 was 2.78 for women compared to 1.97 for men, suggesting that responses among women more sharply distinguished between adjacent levels of RMA.

**Table 4**

*Multi-Group Graded Response Model Parameter Estimates for the “She Asked For It” Subscale by Gender*

Item	$a$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
<b>Men (n = 509)</b>					
1	2.893***	1.248***	1.827***	2.384***	3.214***
2	1.970***	-0.028***	0.759***	1.661***	2.666***
3	2.723***	0.398***	0.988***	1.747***	2.455***
4	1.617***	-0.256	0.624***	1.906***	2.862***
5	2.454***	0.892***	1.565***	2.467***	2.879***
6	1.656***	-0.602***	0.240	1.207***	2.387***
<b>Women (n = 494)</b>					
1	3.324***	1.354***	1.744***	2.471***	2.808***
2	2.779***	0.794***	1.233***	1.807***	2.634***
3	3.165***	0.826***	1.300***	1.930***	2.410***
4	1.954***	0.250***	0.982***	2.072***	2.771***
5	2.393***	1.398***	1.972***	2.912***	3.198***
6	1.894***	0.415***	1.082***	1.786***	2.508***

*Note.* Item discrimination ( $a$ ) and category difficulty parameters ( $\beta_1$ – $\beta_4$ ) were estimated from the multi-group GRM. Higher  $a$ -values indicate greater item discrimination. Category difficulty parameters represent locations along the latent trait continuum at which respondents have a 50% probability of endorsing a given response category or higher (i.e., cumulative category probabilities).

\*\*\* $p < .001$ .

Category difficulty parameters ( $\beta_1$ – $\beta_4$ ) were consistently lower for men across Items 2, 3, 4, and 6 (see Table 4), reflecting the uniform DIF pattern identified in Table 3. Item 1 showed no meaningful differences in either discrimination or category difficulty parameters, supporting invariance at the item level.

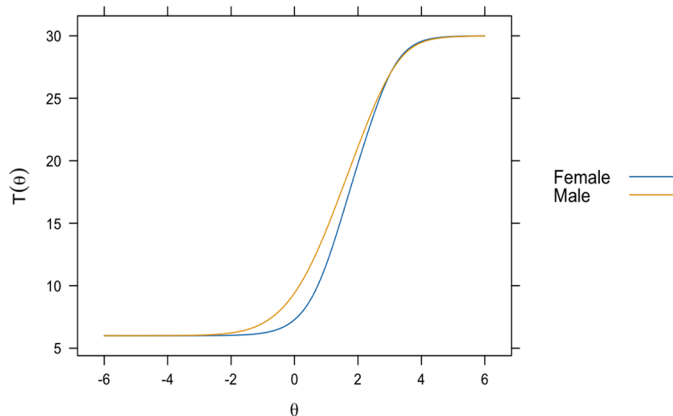
## Scale-Level Impact of Differential Item Functioning

To assess the cumulative impact of item-level DIF on total subscale scores, DRF analyses were conducted (Chalmers, 2018). DRF estimates whether expected total scores differ between groups at equivalent levels of the latent trait, thereby quantifying scale-level score differences attributable to item-level DIF.

Figure 2a displays expected total scores across the latent trait continuum for men and women, whereas Figure 2b presents the corresponding differences in expected total scores between groups, along with bootstrap confidence intervals. Conditional on equivalent levels of latent RMA, men obtained higher expected total scores than women across most trait levels. At the latent trait mean ( $\theta = 0$ ), the expected total score was 9.42 for men and 7.28 for women, reflecting a difference of 2.14 points despite equivalent latent trait levels. The maximum observed difference was 2.85 points at  $\theta = 0.8$  (slightly above the population mean). Scale-level differences were substantially smaller at very low ( $\theta < -2$ ) and very high ( $\theta > +3$ ) trait levels, where floor and ceiling effects reduce variability in expected scores. Notably, DRF effects were concentrated in the central region of the latent trait continuum (approximately  $\theta = -1$  to  $+1$ ), where most respondents are located.

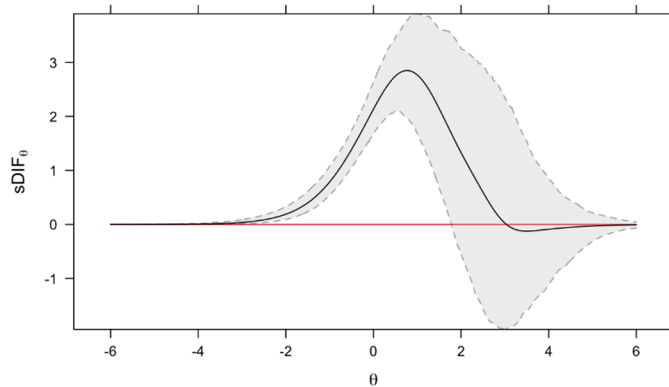
**Figure 2a**

*Expected Total Scale Scores by Gender Across the Latent Trait Continuum*



**Figure 2b**

*Differences in Expected Total Scale Scores Between Men and Women Across the Latent Trait Continuum*



*Note.* The solid black line represents the point estimate of the male–female difference at each level of the latent trait ( $\theta$ ); the shaded grey band represents the 95% confidence interval derived from 500 bootstrap draws. The red horizontal reference line indicates no difference. Positive values indicate higher expected total scores for men than women at equivalent levels of latent rape myth acceptance.

Although five of the six items exhibited statistically significant DIF, the cumulative impact at the scale level was modest in practical terms. Across the central region of the latent trait continuum, differences corresponded to approximately 2–3 points on the 30-point scale, representing a relatively small proportion of the total score range. Bootstrap confidence intervals (see Figure 2b) indicate that this scale-level difference is statistically distinguishable from zero across the central region of the latent continuum (approximately  $\theta = -1$  to  $+2$ ), where the majority of respondents are located, while becoming increasingly uncertain at higher trait levels, where ceiling effects and reduced sample density limit precision. These findings indicate that statistically significant item-level DIF does not necessarily produce large distortions in total scores, and that the practical impact of DIF should be evaluated in terms of magnitude, precision, and distribution across the latent continuum rather than statistical significance of DIF tests alone.

## Discussion

This study evaluated the psychometric performance of a Spanish adaptation of the IRMA-S “She Asked For It” subscale, with a focus on gender-based measurement invariance. The findings provide a nuanced account of comparability across gender groups. At the subscale level, MG-CFA analyses supported metric invariance, indicating that item thresholds and factor loadings operated similarly across men and women. These results

suggest that the latent construct is represented in a comparable way across gender groups and that comparisons of latent means are statistically defensible.

However, it should be noted that metric invariance was established in a re-specified model that allowed residuals of Items 1 and 5 to correlate freely across both groups. This adjustment followed a high modification index in the male subgroup, reflecting local item dependence between two items sharing closely related victim-blaming content (*i.e.*, attributions of responsibility based on intoxication and unclear refusal). The stronger residual correlation observed for men than for women may suggest that these beliefs are more tightly linked in men’s cognitive representations of rape myths, a pattern that warrants further investigation in future studies.

Item-level analyses revealed that several items exhibited differential functioning across gender, with the detected DIF being predominantly uniform in nature. These results highlight the importance of complementing scale-level invariance testing with item-level evaluation. As noted in methodological research on measurement invariance, scale-level equivalence does not necessarily preclude the presence of differential functioning at the item level (Zumbo, 2003). Importantly, the scale-level impact of these differences was modest. Although DIF was detected in most items, DRF analyses indicated that the resulting differences in expected total scores between men and women were relatively small, even in the central region of the latent trait continuum where they were most pronounced. Moreover, bootstrap confidence intervals indicated that these scale-level differences were most precisely estimated within the central region of the latent trait continuum, where respondent density was highest. This pattern illustrates that statistically detectable item-level non-invariance does not necessarily imply substantial distortions in total scores. It also underscores the importance of considering the practical magnitude of DIF when interpreting measurement comparability.

## The Value of Multi-Method Invariance Testing

A central contribution of this study is methodological. The combined application of MG-CFA, item-level DIF analysis, and scale-level DRF assessment illustrates the value of triangulating approaches to measurement invariance. Rather than providing a single binary judgment of equivalence, these methods offer complementary perspectives on different layers of cross-group comparability. Subscale-level invariance testing evaluates whether the latent construct is represented similarly across groups, whereas item-level analyses identify potential parameter-level differences in item functioning. DRF analyses, in turn, provide an assessment of the practical impact of such differences on expected scale scores (Millsap, 2011; Zumbo et al., 2007). Considering these levels jointly allows researchers to distinguish between statistically detectable item-level non-invariance and its substantive implications for scale interpretation.

This distinction is particularly relevant in the study of RMA, where previous cross-cultural validations of IRMA-based instruments have often relied primarily on CFA-based

tests of invariance (e.g., Balezina & Zakharova, 2023; Nyúl & Kende, 2023). Although such analyses are essential for establishing structural comparability, they may not fully capture parameter-level differences in item functioning. Indeed, some studies have reported mixed patterns of invariance across items, even when overall scale-level comparability appeared adequate (Fakunmoju et al., 2019).

More broadly, methodological research has emphasized that statistically significant DIF does not necessarily translate into meaningful distortions in scale-level scores, underscoring the importance of evaluating both item-level functioning and its cumulative impact at the scale level (Zumbo, 2003). Integrating CFA-based invariance testing with IRT-based DIF and DRF analyses provides a more comprehensive framework for assessing cross-group comparability. Researchers conducting cross-group validation studies are encouraged to adopt multi-method strategies to ensure robust and substantively interpretable conclusions.

## Interpreting Item-Level Differences

From a substantive perspective, the detected DIF was concentrated in items addressing beliefs about female sexuality, clothing, and sexual signaling cues—elements that are often central to victim-blaming narratives in rape myth discourse. In particular, Items 2 and 6 reflect cultural scripts linking women's appearance or sexual behavior to perceived responsibility for sexual victimization, themes widely identified as core components of RMA (Burt, 1980; Suarez & Gadalla, 2010). At equivalent levels of the latent trait, men showed a higher probability than women of endorsing higher response categories on these items, primarily due to lower category difficulty parameters. This pattern was most pronounced in the moderate range of RMA ( $\theta$  between  $-1$  and  $+1$ ), where the majority of respondents are located, suggesting that the impact of DIF is conditional on respondents' position along the latent trait continuum rather than homogeneous across it.

The magnitude of these effects suggests that some previously reported gender differences in raw total scores may be partially influenced by measurement-related factors. Differences of approximately 2–3 points in observed scores may reflect, at least in part, item-level differential functioning in addition to substantive attitudinal variation. Although this does not call into question the robust evidence of gender disparities in RMA—meta-analytic research has documented moderate to large gender differences, with men consistently reporting higher levels of RMA than women (Suarez & Gadalla, 2010)—it underscores the importance of distinguishing between true differences at the latent construct level and distortions introduced at the measurement level.

Importantly, the scale-level impact of DIF, as assessed through DRF, was relatively small. These findings indicate that the translated subscale remains appropriate for research applications, particularly within latent variable modeling frameworks that explicitly model measurement parameters and group differences. The presence of item-level non-invariance does not invalidate the instrument; rather, it calls for cautious interpre-

tation of raw total score comparisons between men and women and highlights the importance of analytic strategies that account for potential measurement non-invariance.

## Practical Implications for Researchers

Several practical implications follow from these findings. When comparing gender groups using the “She Asked For It” subscale, differences in raw total scores should be interpreted with caution, particularly when observed differences are relatively small in magnitude. DRF analyses indicate that such differences may partially reflect item-level differential functioning in addition to substantive variation in the underlying construct. These findings suggest that researchers comparing men and women should avoid relying exclusively on raw total scores when drawing substantive conclusions about gender differences in RMA. Latent variable modeling approaches, which explicitly model measurement parameters and potential non-invariance, provide a more robust framework for cross-group comparisons (Putnick & Bornstein, 2016).

The findings further reinforce that DIF should not be interpreted as a purely binary phenomenon. Statistical significance alone is insufficient to determine substantive distortion; the magnitude of parameter differences and their distribution along the latent trait continuum must also be considered (Millsap, 2011; Zumbo et al., 2007). In the current analyses, DIF effects were concentrated primarily within the moderate range of the trait, indicating that potential bias varies according to respondents’ underlying level of RMA rather than affecting all individuals uniformly.

## Cultural and Context Considerations

The pronounced DIF observed in items related to clothing, sexual signaling cues, and ambiguous consent may reflect gender-related differences in how such situations are interpreted within contemporary Spanish discourse. Legal reforms and public debate following the *La Manada* case and the 2022 “Solo Sí es Sí” law have reshaped societal understandings of consent and victim responsibility (Larrauri, 2020; Villacampa & Pujols, 2022). These developments may help explain the pronounced floor effects observed among women for more explicit victim-blaming statements, whereas more situationally or normatively framed items appear to allow greater interpretative variability. Accordingly, DIF effects were especially pronounced in items requiring evaluative judgements about women’s behavior.

That DIF effects were most pronounced in the moderate range of the trait indicates that measurement differences are not uniformly distributed but instead vary as a function of latent endorsement levels. Future research could examine potential cognitive processing differences across gender using qualitative approaches such as cognitive interviewing and think-aloud protocols. Extending invariance testing to additional sociodemographic and linguistic groups would further clarify whether observed item-level

non-invariance reflects translation nuances, contextual framing effects, or broader differences in the interpretation of contested social norms (Putnick & Bornstein, 2016).

## Limitations and Future Directions

Several limitations of the current study warrant consideration. First, the sample was drawn exclusively from one region in Spain (Castilla-La Mancha), which may limit the generalizability of the findings to the broader Spanish population. Although quota sampling ensured balance by age and gender, other potentially influential variables—such as education level or political orientation—were not adjusted and may have influenced response patterns. Notably, the sample differed from the Castilla-La Mancha population in educational attainment, with a higher proportion of respondents holding a college degree (41.8%) compared to the target population (25.9%).

Second, while the IRMA-S remains one of the most widely used instruments for assessing RMA, the subscale analyzed here captures only a specific domain: victim-blaming attitudes based on perceived provocation. As public discourse and social norms evolve, new and more nuanced forms of rape myths may emerge. Future research should examine the psychometric performance and measurement invariance of these subtler forms of RMA across diverse demographic and cultural contexts.

Third, the decision to include only one of the four IRMA-S subscales was guided by theoretical relevance and practical constraints. This necessarily limits the scope of the construct assessed and restricts the generalizability of the findings. Because the remaining subscales were not administered, conclusions about the measurement properties of the full instrument cannot be drawn. Future research should validate the complete IRMA-S in Spain to explore whether similar patterns of item-level non-invariance and scale-level impact are observed across other dimensions of rape myths, such as rape denial or perpetrator exoneration.

Fourth, as this study focused solely on the Spanish adaptation of the “She Asked For It” subscale from the IRMA-S, cross-language measurement invariance could not be evaluated. Although the translation followed a structured TRAPD process and aimed to ensure conceptual equivalence, future studies should directly test language-based invariance between the English and Spanish versions to confirm that both operate equivalently across linguistic and cultural contexts.

Finally, the pronounced floor effects observed—particularly among women—reduced response variability, which may have influenced the estimation of item parameters and contributed to the detection of item-level DIF. Although DRF analyses indicated that the cumulative scale-level impact of DIF was modest, restricted variability may still affect the precision with which moderate levels of the latent trait are measured. These findings highlight the importance of continued psychometric evaluation of RMA instruments to ensure sensitivity across the full range of the construct. Future scale development should prioritize items capable of capturing subtle and implicit attitudes, especially in

populations where overt endorsement of explicit rape myths has become increasingly rare.

## Conclusion

This study highlights the importance of rigorously evaluating the Spanish adaptation of the IRMA-S “She Asked For It” subscale presented in this study before using it for gender-based comparisons in RMA. Although the subscale demonstrated strong internal consistency, a stable unidimensional structure, and metric invariance at the subscale level, item-level analyses conducted within a multi-group IRT framework revealed DIF in most items. Although the cumulative scale-level impact of DIF was modest, these findings indicate that observed gender differences in raw scores may partly reflect measurement artifacts in addition to substantive differences in RMA.

More broadly, the findings underscore the need for rigorous and multi-layered validation when adapting attitudinal instruments for comparative or policy-oriented research. Instruments designed to assess sensitive constructs such as RMA must be both culturally appropriate and psychometrically sound. In this regard, measurement invariance should not be treated as a binary property of a scale but as a multi-layered construct that requires evaluation at the item and scale levels. Careful examination of these layers is therefore not merely a technical procedure but a fundamental requirement for ensuring valid and interpretable conclusions about group differences in both academic and applied contexts.

---

**Funding:** This research was funded by the Institute for Women of Castilla-La Mancha, Research Grants in 2023 [2023/7446].

---

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their thoughtful comments and suggestions on earlier versions of this manuscript. We are especially grateful to the handling editor, Piotr Koc, whose insightful feedback and guidance greatly helped to improve the paper.

---

**Competing Interests:** The authors have declared that no competing interests exist.

---

**Data Availability:** The data are available from the corresponding author upon reasonable request.

---

## Supplementary Materials

For this article, the following Supplementary Materials are available (see [Leon et al., 2026](#)):

- Table 1A. Comparison between the study sample and the population of Castilla-La Mancha in terms of gender, age, education, and political orientation.
- Structured Review Template for Expert Pretesting.

- Template used by the expert panel to assess the semantic clarity and interpretability of the survey items.
- Table 2A. Original English wording and Spanish-translated wording of the items included in the “She Asked For It” subscale of the Illinois Rape Myth Acceptance Scale – Subtle Version (IRMA-S).

### Index of Supplementary Materials

Leon, C. M., Aizpurua, E., & Quiñonez-Toral, T. (2026). *Supplementary materials to “Gender dynamics in rape myth acceptance: A psychometric evaluation of the IRMA-S “She asked for it” subscale in Spanish”* [Tables, Templates]. PsychOpen GOLD.  
<https://doi.org/10.23668/psycharchives.22166>

## References

- Balezina, M., & Zakharova, S. (2023). Measuring attitudes towards rape in Russia: Translation and validation of the Illinois Rape Myths Acceptance scale. *Current Psychology*, 43(5), 4611–4621.  
<https://doi.org/10.1007/s12144-023-04666-2>
- 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>
- Chalmers, R. P. (2018). Model-based measures for detecting and quantifying response bias. *Psychometrika*, 83(3), 696–732. <https://doi.org/10.1007/s11336-018-9626-9>
- Delegación del Gobierno Contra la Violencia de Género. (2017). *Percepción social de la violencia sexual*.  
[http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/3180\\_3199/3182/cues3182.pdf](http://www.cis.es/cis/export/sites/default/-Archivos/Marginales/3180_3199/3182/cues3182.pdf)
- Dong, Y., & Dumas, D. (2020). Are personality measures valid for different populations? A systematic review of measurement invariance across cultures, gender, and age. *Personality and Individual Differences*, 160, Article 109956. <https://doi.org/10.1016/j.paid.2020.109956>
- D’Urso, E. D., De Roover, K., Vermunt, J. K., & Tilmstra, J. (2022). Scale length does matter: Recommendations for measurement invariance testing with Categorical Factor Analysis and Item Response Theory approaches. *Behavior Research Methods*, 54(5), 2114–2145.  
<https://doi.org/10.3758/s13428-021-01690-7>
- Fakunmoju, S. B., Abrefa-Gyan, T., & Maphosa, N. (2019). Confirmatory factor analysis and gender invariance of the revised IRMA scale in Nigeria. *Affilia*, 34(1), 83–98.  
<https://doi.org/10.1177/0886109918803645>
- Fansher, A. K., & Zedaker, S. B. (2022). The relationship between rape myth acceptance and sexual behaviors. *Journal of Interpersonal Violence*, 37(1–2), 903–924.  
<https://doi.org/10.1177/0886260520916831>
- Genc, A., Kopilović, D., & Dinić, B. M. (2026). A psychometric evaluation of the Illinois Rape Myth Acceptance Scale – subtle version (IRMA-S-2022) and the newly developed Rape Myth

- Questionnaire – short form (RMQ-S) in Serbian culture. *Sexual Abuse: A Journal of Research and Treatment*, 38(2), 163–186. <https://doi.org/10.1177/10790632251334755>
- González-Caino, P. C., Resett, S., Lopez, J. I., & Bossi, F. (2022). El cuestionario sobre mitos de violación: Propiedades psicométricas, psicopatía y autoestima. *Revista Ecuatoriana De Psicología*, 5(13), 198–213. <https://doi.org/10.33996/repsi.v5i13.82>
- Kazmi, S. M. A., Farooq, Z., & Tariq, S. (2024). Adaptation of the updated Illinois Rape Myth Acceptance Scale in Urdu. *Sexuality & Culture*, 28(4), 1496–1511. <https://doi.org/10.1007/s12119-023-10189-6>
- Kleinsmith, O. M., Hahn, O. N., Henry, D., Merrell, L., Blackstone, S., Tomchik, A. Y., Sell, J., Gramstad, M., Schuetz, J. E., Doyle, C. M., & Haney, B. M. (2026). Are rape myths inherently gendered? Examining assumed gender ascribed to gender-neutral versions of the Illinois Rape Myth Acceptance scale among college students. *Violence Against Women*, 32(5), 1256–1270. <https://doi.org/10.1177/10778012251329222>
- Larrauri, E. (2020). Criminalizing sexual violence in Spain: From “abuse” to “assault” and the debate on consent. *New Criminal Law Review*, 23(3), 320–338.
- Leon, C. M., Aizpurua, E., Quiñonez-Toral, T., & Rollero, C. (2025). Understanding rape myth acceptance through the lens of sexual objectification theory: The role of pornography consumption, purchase of sexual services, and masculinity. *Journal of Sex Research*, 62(9), 1892–1904. <https://doi.org/10.1080/00224499.2024.2446635>
- Martini, M., Tartaglia, S., & De Piccoli, N. (2022). Assessing Rape Myth acceptance: A contribution to Italian validation of the Measure for Assessing Subtle Rape Myth (SRMA-IT). *Sexual Abuse: A Journal of Research and Treatment*, 34(4), 375–397. <https://doi.org/10.1177/10790632211028158>
- McMahon, S., & Farmer, G. L. (2011). An updated measure for assessing subtle rape myths. *Social Work Research*, 35(2), 71–81. <https://doi.org/10.1093/swr/35.2.71>
- Megías, J. L., Romero-Sánchez, M., Durán, M., Moya, M., & Bohner, G. (2011). Spanish validation of the Acceptance of Modern Myths about Sexual Aggression Scale (AMMSA). *The Spanish Journal of Psychology*, 14(2), 912–925. [https://doi.org/10.5209/rev\\_SJOP.2011.v14.n2.37](https://doi.org/10.5209/rev_SJOP.2011.v14.n2.37)
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.
- Nyúl, B., & Kende, A. (2023). Rape myth acceptance as a relevant psychological construct in a gender-unequal context: The Hungarian adaptation of the updated Illinois Rape Myth Acceptance scale. *Current Psychology*, 42(4), 3098–3111. <https://doi.org/10.1007/s12144-021-01631-9>
- Payne, D. L., Lonsway, K. A., & Fitzgerald, L. F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois Rape Myth Acceptance scale. *Journal of Research in Personality*, 33(1), 27–68. <https://doi.org/10.1006/jrpe.1998.2238>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Suarez, E., & Gadalla, T. M. (2010). Stop blaming the victim: A meta-analysis on rape myths. *Journal of Interpersonal Violence*, 25(11), 2010–2035. <https://doi.org/10.1177/0886260509354503>

- Svetina, D., Rutkowski, L., & Rutkowski, D. (2019). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/semTools packages. *Structural Equation Modeling*, *27*(1), 111–130.  
<https://doi.org/10.1080/10705511.2019.1602776>
- Tay, L., Meade, A. W., & Cao, M. (2014). An overview and practical guide to IRT measurement Equivalence analysis. *Organizational Research Methods*, *18*(1), 3–46.  
<https://doi.org/10.1177/1094428114553062>
- Thelan, A. R., & Meadows, E. A. (2021). The Illinois Rape Myth Acceptance Scale—Subtle version: Using an adapted measure to understand the declining rates of rape myth acceptance. *Journal of Interpersonal Violence*, *37*(19–20), NP17807–NP17833.  
<https://doi.org/10.1177/08862605211030013>
- Villacampa, C., & Pujols, A. (2022). The reform of sexual offences in Spain: Affirmative consent and the impact of the “Only Yes Means Yes” law. *International Journal of Law, Crime and Justice*, *70*, Article 100540.
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045.  
<https://doi.org/10.1007/s11336-016-9506-0>
- Zumbo, B. D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language Testing*, *20*(2), 136–147.  
<https://doi.org/10.1191/0265532203lt248oa>
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21–29.  
<https://doi.org/10.22237/jmasm/1177992180>