Evaluating Attitudes Toward Sexual and Gender Diversity in Education: A Systematic Review of Assessment Tools

Daniel Serey
Universidad San Sebastián. Chile
https://orcid.org/0000-0001-6236-836X

María Aranda
Universidad de Jaén
https://orcid.org/0000-0003-1905-8379

Samuel P. León
Universidad de Jaén
https://orcid.org/0000-0002-6980-2680

The inclusion of sexual and gender diversity in education remains a persistent challenge, largely due to the enduring presence of negative attitudes. Education is a key context in which prejudice can be reinforced or, alternatively, egalitarian attitudes promoted. Advancing inclusion requires valid instruments to assess attitudes toward diversity. To this end, several authors have developed, validated, or applied instruments to assess attitudes toward sexual and gender diversity in education. The aim of this study was to systematically review these instruments. Specifically, we pursued three objectives: (1) to identify and analyze the constructs targeted by these measures, examining their adequacy for capturing both explicit and subtle attitudes (RQ1); (2) to synthesize the descriptive characteristics of the studies in which they were applied, including information on target populations, educational levels, and cultural or geographical contexts (RQ2); and (3) to evaluate the methodological quality of scale development and the psychometric properties reported, in line with international standards (RQ3). We conducted a systematic review with meta-analyses, following the recommendations of PRISMA and APA standards. After the eligibility criteria were applied, seven studies were included. Results indicate that most scales provide some evidence of validity and reliability. To reduce social desirability, some instruments combine positively and negatively worded items and incorporate indicators of modern prejudice. However, certain items may no longer adequately capture subtle forms of bias or reflect evolving terminology. Future research should address challenges such as the over-generality of items, the limitations of outdated prejudice measures, and the need for stronger evidence to support psychometric properties.

***Keywords**:* Education; Attitudes; Sexual diversity; Gender identity; Measures.

In recent decades, numerous initiatives have sought to advance equity in the exercise of human rights for LGBTIQ+ people, with particular emphasis on self-determination, identity, and expression (Alegre & Fiedler, 2021; Maschi et al., 2022). Yet, under the influence of political, social, and religious factors, LGBTIQ+ rights still vary widely across regions and countries (European Commission, 2023; European Union Agency for Fundamental Rights [FRA], 2024; Mella Gómez, 2024). Even in parts of the world that have made significant progress in human rights, paradoxes persist: while more LGBTIQ+ people are openly expressing their identities, they simultaneously face higher levels of violence, harassment, and intimidation, particularly among younger individuals (FRA, 2024; McKay et al., 2017).

Educational institutions reflect these dual dynamics. While such environments can perpetuate violence and discrimination through physical aggression, verbal abuse, cyberbullying, exclusion, and isolation (Özdere, 2023), they also have the potential to serve as safe and inclusive spaces for personal and social development (Epps et al., 2023). Research highlights the presence of inclusive practices within the education system but also points to significant gaps, such as ignorance, stereotypes, and taboos—that hinder the effective integration of sexual and gender diversity (Bartual-Figueras et al., 2023; Carrión et al., 2020). These dynamics shape peer and teacher–student relationships and have serious consequences for academic performance, well-being, and long-term health (Moral et al., 2025). International organizations stress that discrimination and violence in schools contribute to elevated levels of fear, anxiety, depression, absenteeism, and dropout among

Corresponding author: Samuel P. León, Campus de las Lagunillas, SN, CP 23071 Jaén, Spain. sparra@ujaen.es

LGBTIQ+ students (Organisation for Economic Co-operation and Development [OECD], 2023; Suarez et al., 2024; UNESCO, 2019).

Addressing these challenges requires identifying and transforming the attitudes that sustain discrimination (Compare et al., 2024; Dovidio et al., 2025). In educational contexts, this involves creating safe environments that promote respect for diversity, establishing clear policies against bullying and harassment, and providing training for teachers, staff, and students (Özdere, 2023). To effectively implement such strategies, it is necessary to assess existing attitudes toward sexual and gender diversity. Such attitudes can be understood as psychological evaluations integrating cognitive, affective, and behavioral components (Ajzen, 2001), directed toward individuals or groups with diverse sexual orientations or gender identities, that is, the spectrum of identities, orientations, and expressions beyond heteronormative and cisnormative assumptions, including lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other non-binary identities (UNESCO, 2020). Reliable and valid instruments are essential for both research and educational practice, as they enable the identification of prevailing biases and stereotypes that hinder inclusion (Kosciw et al., 2022).

Systematic scholarly investigation of attitudes toward sexual and gender diversity began to attract greater attention in the 1970s. One of the first recognized tools was the Attitudes Toward Homosexuality Scale (ATHS; Hudson & Ricketts, 1980). Early instruments often reflected pathologizing or moralizing assumptions, but a turning point came with Herek's Attitudes Toward Lesbians and Gay Men Scale (ATLG, 1998), regarded as one of the most influential tools with strong psychometric properties (Lara-Garrido et al., 2024). This instrument differentiates between attitudes toward gay men and lesbians and has been adapted and applied in numerous countries (e.g., Bulboacă et al., 2021; Luu et al., 2025; Wang et al., 2022). Since the 1990s, the use of these instruments in educational contexts has steadily increased. Some studies have adapted general instruments such as the ATHS or ATLG, whereas others have developed and validated scales specifically for educational settings (e.g., Attitudes Toward Lesbians, Gay Men, and Bisexuals Scale; Ensign et al., 2011). Nevertheless, a comprehensive review is still lacking to determine the extent to which these instruments capture contemporary expressions of attitudes toward sexual and gender diversity, the breadth of their application across educational levels and cultural contexts, and the robustness of their methodological and psychometric foundations.

This gap is particularly relevant because attitudes have shifted from overt hostility to more subtle and ambivalent expressions (Dovidio et al., 2025; Herek & McLemore, 2013), even as explicit prejudice persists (Herek, 1988; Morrison & Morrison, 2003). Scholars have also questioned the assumption that psychological measures are universally applicable across identities (Cramwinckel et al., 2018), since instruments developed within heteronormative frameworks risk excluding diverse genders and sexualities (Bravestone et al., 2024; Matsick et al., 2025; Stiekema et al., 2024). Accordingly, it is essential to examine whether instruments used in educational contexts measure traditional or subtle forms of prejudice, whether they assess general or group-specific biases (e.g., homophobia, biphobia, transphobia), and whether they capture constructs such as modern homonegativity, microaggressions, or ostensibly positive but paternalistic behaviors.

These conceptual considerations are particularly salient in educational systems, which can reproduce broader social inequalities through heteronormative and binary gender frameworks, thereby fostering discrimination against LGBTIQ+ students (Shoaei, 2021). Such discrimination manifests differently across groups: students may engage in overt behaviors such as bullying or cyberbullying (Garaigordobil & Larrain, 2020), whereas teachers often display subtler biases, for example by avoiding topics or limiting inclusive practices (Formby, 2011; Payne & Smith, 2018). This underscores the importance of considering target populations and educational levels when evaluating instruments. Cultural and geographical contexts are also critical, as attitudes toward sexual and gender diversity vary across societies, and scales may not function equivalently without evidence of measurement invariance (Gromadzki, 2019; Lara-Garrido et al., 2024; Muñoz-García et al., 2023). Consequently, systematic reviews must synthesize not only the constructs measured but also the characteristics of the studies, such as populations, educational levels, and cultural/geographical settings, to evaluate the generalizability and applicability of instruments (López-Orozco et al., 2022; UNESCO, 2020).

Taken together, these considerations highlight the need to ensure that instruments meet rigorous psychometric standards. Reliable and valid measurement is

particularly critical for attitudes toward sexual and gender diversity, which are dynamic and shaped by cultural, social, and personal factors (Aggarwal et al., 2024; Morrison et al., 2018; Stanick et al., 2021). Key aspects include reliability, consistency across time and settings, and validity, the extent to which the instrument accurately measures the intended construct (Bishop & Pynoo, 2020; DeVellis & Thorpe, 2021). Methodological rigor in scale development is also crucial to ensure transparency and robustness (Swan, 2023). Systematic reviews reveal substantial variability in instrument quality, with many scales assessing LGBTIQ+ prejudice lacking adequate evidence for item development, scale structure, reliability, and validity (Aggarwal et al., 2024; Bishop & Pynoo, 2020; Morrison et al., 2018). Despite ongoing efforts, a notable gap remains regarding instruments applied in educational contexts, where reliable and valid measures are essential to advance both research and interventions.

The present study

The objective of this review was to analyze instruments developed to assess attitudes toward sexual and gender diversity in educational contexts. Specifically, three objectives were established, each linked to a research question:

Objective 1. To identify and analyze the constructs targeted by these instruments. **RQ1**. Which constructs are targeted by existing instruments (e.g., sexual diversity broadly, gender identity only, or sexual-affective orientation only), and how adequate are they for capturing both explicit and subtle attitudes?

Objective 2. To synthesize the descriptive characteristics of the included studies in which these instruments were applied, including information on target populations, educational levels, and cultural or geographical contexts. **RQ2**. What are the descriptive characteristics of the included studies (e.g., population, educational level, geographical/cultural context) in which these instruments were applied?

Objective 3. To evaluate the methodological quality of the scale development processes and the psychometric properties reported, following established frameworks (VALID, COSMIN, APA Standards, DeVellis & Thorpe, 2021). **RQ3**. What is the methodological quality of the development processes, and what psychometric properties (e.g., validity, reliability, factorial structure, measurement invariance) have been reported according to international standards?

1. Methods

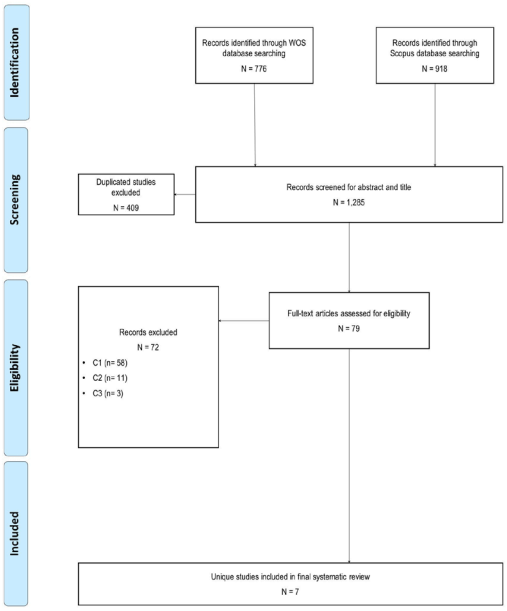
1.1. Search procedures

This systematic review follows the recommendations of PRISMA (Page et al., 2021) and the APA's reporting standards for systematic reviews and meta-analyses (Appelbaum et al., 2018) (for further details of the PRISMA process, see Appendix D).

On April 26, 2024, a search was conducted in the Web of Science (WoS) and Scopus databases using the following search equation in the topical section: ((Attitudes OR Perception) AND (Diversity) AND (Education) AND (Validation OR Scale)). On July 12, 2024, we conducted a second search with the same parameters but using the following equation: ((Attitudes OR Perception) AND ("gender identity" OR "sexual orientation") AND (Education) AND (Validation OR Scale)). The sum of both searches, after eliminating 409 duplicates, yielded a total of 1,285 articles. All studies considered for inclusion had to meet the following criteria: (C1) studies measuring attitudes and perceptions toward sexual diversity and gender identity in an educational context; (C2) studies employing a quantitative methodology; and (C3) studies conducted in an educational context. The search was limited to works published in English, with no temporal restrictions applied. After applying these criteria, seven studies met the eligibility requirements and were retained for the final review, as shown in the PRISMA flow diagram (Figure 1; see Appendix D).

The first author performed the initial screening of the 1,285 articles by examining the titles and abstracts, resulting in 79 articles provisionally included. Subsequently, the full-text assessment of these 79 articles was independently performed by both the first and last authors, applying the inclusion criteria to determine final eligibility (88.60% initial inter-rater agreement; discrepancies were discussed until 100% consensus was reached). As a result, seven studies were included in the final sample for this review (see Table 1). Figure 1 presents the summary of the filtering process through a PRISMA flowchart.

Figure 1. PRISMA flowchart.



1.2. Data extraction and coding

The 7 studies finally included were analyzed in depth by the authors. For each included study, we coded information at three levels: (a) the constructs and dimensions assessed by the instrument; (b) the descriptive characteristics of the study (e.g., sample, educational level, country); and (c) the methodological quality of the scale development process and the psychometric properties reported.

To evaluate the quality of the scales on attitudes and perceptions towards sexual diversity and gender identity included in this systematic review, we assessed both scale development rigor and psychometric properties using three complementary frameworks; a) An adapted version of VALID Checklist (Kerschbaumer et al., 2025). This scale was used to evaluate the general validity and methodological transparency; b) COSMIN/APA Guidelines (Terwee et al., 2018; American Psychological Association [APA] & National Council on Measurement in Education, 2020), to evaluate the adequacy and robustness of the psychometric treatment of the scales; c) DeVellis' Scale Development Framework (DeVellis & Thorpe, 2021), aimed at evaluating the rigor followed by the work in the procedures developed in the construction of the scales (Detailed operationalization of each framework (VALID, COSMIN/APA, DeVellis & Thorpe, 2021) is provided in Appendix A).

1.3. Computation of effect sizes and statistical analyses

To analyze the psychometric properties of the scales analyzed, we extracted the information provided in the studies regarding the reliability indices, as well as the fit indices after performing a factor analysis. In the case of internal consistency, all the alpha coefficients were transformed following Bonett (2002) to allow the normalization of the distribution of the coefficients, as well as the stability of their variances. In order to facilitate the understanding of the results, the resulting coefficients were transformed back to Cronbach's alpha (see Appendix B of the supplementary material).

Given that some of the included studies presented more than one index for the same sample, we fitted a multilevel random effects model to avoid issues related to the independence of the effects using the *rma.mv()* function of the R package 'metafor' (Viechtbauer, 2010). For interpretative purposes, reliability coefficients were evaluated against conventional thresholds: values ≥ .70 were considered acceptable for exploratory purposes, and values ≥ .80 were taken as strong evidence of internal consistency (Nunnally & Bernstein, 1994; Tavakol & Dennick, 2011). All raw data, analysis scripts, and supplementary material have been made publicly available through the OSF repository. (https://osf.io/qrtb5/).

2. Results

2.1. Constructs measured by the instruments

The instruments included in this review varied in the constructs they were designed to measure. Three focused primarily on homophobic attitudes (e.g., the ADAS; Garrido-Hernansaiz et al., 2018; the MHS; Ensign et al., 2011; Barragán & Pérez-Jorge, 2020), whereas one was directed specifically toward attitudes toward transgender individuals (the EANT; Alonso et al., 2021). Two instruments addressed attitudes toward sexual diversity more broadly, without differentiating between sexual orientation and gender identity (e.g., Morales et al., 2020;

Ospina-Betancurt et al., 2023). Finally, Ensign et al. (2018) adapted an existing scale to explore teachers' attitudes toward supporting LGBTIQ+ students, again focusing on overt homophobic attitudes.

In terms of how attitudes were conceptualized, most instruments adopted a rather narrow approach, focusing on explicit evaluative judgments captured through Likert-type items. Only a minority explicitly addressed the multidimensional nature of attitudes (cognitive, affective, and behavioral components), and none employed indirect measures capable of capturing implicit bias.

Importantly, several constructs that are central in the literature on sexual and gender diversity were entirely absent. None of the included instruments assessed binegativity, despite evidence of its unique role in shaping prejudice toward bisexual populations (Herek & McLemore, 2013). Similarly, microaggressions, which are increasingly recognized as a subtle but pervasive form of discrimination (Sue, 2010), were not covered. Moreover, while modern forms of prejudice (e.g., benevolent or symbolic homophobia) have been documented in the literature, they were scarcely reflected in the instruments identified here.

In summary, the instruments reviewed mainly capture explicit homophobic attitudes and, to a lesser extent, transphobic attitudes. However, they rarely consider the multidimensional nature of attitudes, overlook binegativity and microaggressions, and pay little attention to subtle or modern forms of prejudice. Together, these findings address RQ1 by showing that existing measures still focus largely on overt attitudes, while more nuanced dimensions remain insufficiently represented.

2.2. Descriptive characteristics of the included studies

The descriptive characteristics of the seven studies included in this review are summarized in Table 1. Most of the research was conducted in Spain (five studies), with two additional studies carried out in the United States (Ensign et al., 2011; Ensign et al., 2018). This concentration of research in the Spanish context underscores the limited geographical diversity in the current evidence base.

Regarding the populations studied, the majority of instruments were administered to students, particularly at the secondary education level (e.g., Garrido-Hernansaiz et al., 2018; Morales et al., 2020). Two studies targeted university students (Barragán & Pérez-Jorge, 2020; Ospina-Betancurt et al., 2023), while only one involved teachers and school staff (Ensign et al., 2018), highlighting the scarcity of tools validated for educators. Sample sizes varied considerably, ranging from relatively small validation samples of around 170 participants (Barragán & Pérez-Jorge, 2020) to larger surveys with more than 2,800 participants (Morales et al., 2020). In terms of age, the samples included both adolescents (~14–18 years old in secondary education) and young adults in higher education (up to ~34 years old). Gender was reported in binary terms (male/female) in almost all studies, with a single exception (Alonso et al., 2021), which included a non-binary category. This lack of inclusivity in demographic reporting reflects a limitation of the primary studies and aligns with broader critiques of binary gender frameworks in educational research. The instruments themselves also varied in scope and structure. Some were unidimensional scales (e.g., ADAS), while others adopted a multidimensional structure (e.g., EANT with its subscales). The number of items ranged from 9 to 30, reflecting different levels of depth and comprehensiveness. Moreover, the visibility of these instruments in the academic literature was uneven: citation counts ranged from just over 10 citations (Barragán & Pérez-Jorge, 2020) to more than 90 citations (Ensign et al., 2011).

Taken together, the included studies show that most instruments have been tested with student populations in Spain and the United States, using convenience samples and predominantly binary demographic categories. This narrow scope restricts the generalizability of findings and highlights the need for future validations in more diverse cultural settings and with more inclusive demographic measures.

2.3. Methodological quality and psychometric properties

2.3.1. Quality of scale development and validation

The evaluation of the methodological quality of the scale development and validation processes, conducted with reference to the VALID, COSMIN, APA Standards, and DeVellis & Thorpe's Scale Development Framework, is summarized in Table 2. Full details are provided in Appendix C of the Supplementary Materials.

Table 1.
Summary of the included studies.

Author (year)	Journal	Country of the study	Sample	Population	Age	Sex	Educational level	Scale	What does it measure? (Construction)	Factors measured by the scale	Nº items	Nº Cites
Alonso et al (2021)	Sustainability	UK.	362	University students	*M*=21.43 (*SD*=3.42)	Cisgender women: 63% (228). Cisgender men: 32.9% (119). Non-binary gender: 3% (11). Preferred not to say: 1.1% (4)	Higher education (university)	The Scale of Negative Attitudes toward Transgender people (EANT)	Negative attitudes toward transgender people.	Unifactorial	9	15
Barragán & Pérez-Jorge (2020)	Heliyon	Spain (with participation of schools in Germany, Denmark, Italy and Mexico)	Students: 2,847 (evaluated directly) Teachers: 143	Students and teachers in primary and secondary schools	NA	NA	Primary and secondary education.	NA	Attitudes towards sexual diversity, homophobia, lesbophobia, biphobia and transphobia.	NA	NA	25
Ensign et al (2011)	Journal of Athletic Training	USA	964 participants (heterosexuals only, out of a total of 1117 respondents)	Athletic trainers (ATs) employed by National Collegiate Athletic Association institutions (NCAA)	M=32.85 (± 8.5)	49.7% men (478), 50.3% women (486)	Athletic trainers in the context of higher education institutions (NCAA)	Attitudes Toward Lesbians, Gay Men, and Bisexuals Scale (ATLGB)	Attitudes toward lesbians, gay men and bisexuals (LGB)	Attitudes Toward Lesbians (ATL). Attitudes Toward Gay Men (ATG). Attitudes Toward Bisexuals (ATB).	30 items (10 for each subscale): ATL, ATG, ATB).	91

Author (year)	Journal	Country of the study	Sample	Population	Age	Sex	Educational level	Scale	What does it measure? (Construction)	Factors measured by the scale	Nº items	Nº Cites
Ensign et al (2018)	Journal of Athletic Training	USA	Phase 1 (interviews): 6 TAs. Phase 2 (item reduction): 507 TAs (17% response rate). Phase 3 (validation): 393 TAs (13% response rate).	Athletic trainers (ATs) and students of professional and post-professional athletic training programs.	M=34.25 (SD=10.19)	62.3% women (245), 37.7% men (148)	Undergraduate students in athletic training programs, as well as practicing athletic trainers.	Attitudes Toward Transgender Patients (ATTP)	Attitudes of athletic trainers toward transgender patients, assessing affective, cognitive and behavioral aspects.	Clinician education (3 items). Transgender sports participation (3 items). Clinician comfort (4 items).	Total 10 -Clinician education: 3 items - Transgender sport participation: 3 items -Clinician comfort: 4 items	11
Garrido-Hernansaiz et al (2018)	Journal of Homosexuality	Spain	696 students, of whom 676 completed the scale	Secondary school students (8th to 10th grade, ESO in the Spanish education system)	*M*=14.48 (*SD*=1.17)	50.9% men (344), 49.1% women (332)	High school students	Attitudes Toward Affective-Sexual Diversity scale (ADAS)	Attitudes towards affective-sexual diversity (homophobia, lesbophobia, biphobia)	Unidimensional	27	11

Author (year)	Journal	Country of the study	Sample	Population	Age	Sex	Educational level	Scale	What does it measure? (Construction)	Factors measured by the scale	Nº items	Nº Cites
Morales et al (2020)	International Journal of Environmental Research and Public Health	Spain	170	University students of primary education.	*M*=20.86 (18-26 years)	67% female, 33% male	University degree (degree in primary education).	Scale of Attitudes toward Coeducation and the Construction of a New Gender Culture (Adapted from School Doing Gender/Students). Modern Homophobia Scale	Attitudes towards coeducation and gender equality. Attitudes towards homosexuality (homophobia towards gays and lesbians)	1-Attitudes Toward Coeducation Scale: Socio-cultural level. Relational level. Personal level. 2-Modern Homophobia Scale: 2.1-Homophobia toward gays (MHS-G): Personal discomfort. Deviance/Changes. Institutional homophobia. 2.2-Homophobia toward lesbians (MHS-L): Personal discomfort. Deviance/Changes. Institutional homophobia.	1-Attitudes towards coeducation scale (sociocultural, relational, personal): No. of items: NA 2-Modern Homophobia Scale: Homophobia towards gays (MHS-G) subscale: 22 items. Subscale homophobia towards lesbians (MHS-L): 24 items.	43

Author (year)	Journal	Country of the study	Sample	Population	Age	Sex	Educational level	Scale	What does it measure? (Construction)	Factors measured by the scale	Nº items	Nº Cites
Ospina-Betancurt et al (2023)	Sexuality & Culture	Spain	610	Physical Activity and Sport Sciences undergraduate students (PASS)	*M*=21.72 (*SD*=4.12).	68.85% men (420) 31.15% women (190).	University students in Physical Activity and Sports Science programs.	Scale of Attitudes Toward Sexual Diversity Among Athletes	Attitudes towards sexual diversity in sport, including lesbian, gay, bisexual and transgender people	Attitudes Cognitive attitudes towards diversity. Attitudes towards gender stereotypes. Attitudes towards transgression. Affective attitudes towards diversity.	Total 18 Cognitive attitudes towards diversity (5 items). Attitudes towards gender stereotypes (4 items). Attitudes towards transgression (3 items). Affective attitudes towards diversity (6 items).	1

Note. M: Mean; *SD*: Standard Deviation. The number of citations was consulted in Goglee Scholar (date 06/18/2025).

Table 2.

Summary of evaluation results of the scales on attitudes and perceptions towards sexual diversity and gender identity according to VALID, COSMIN/APA, and DeVellis & Thorpe (2021) criteria.

Author (Year)	VALID		COSMIN/APA		DeVellis & Thorpe (2021)		Key limitations
Alonso et al. (2021	Construct	⊕	Reliability	⊕	Steps	6/8	No test-retest or explicit cognitive pilot is reported. Some items could be sensitive (it is suggested to reformulate them in positive).
	Pilot	?	Construct validity	⊕	Pilot	?	
	Items	⊕	Convergent	⊕	Rigorous analysis	⊕	
Barragán et al. (2020)	Construct	⊕	Reliability	⊖	Steps	3/8	No reliability, construct validity or temporal stability data are reported. It is unclear whether pretests were conducted with the target population.
	Pilot	?	Construct validity	⊕	Pilot	?	
	Items	⊕	Convergent	⊖	Rigorous analysis	⊖	
Ensign et al. (2011)	Construct	⊕	Reliability	⊕	Steps	4/8	Evidence of construct validity (CFA), convergent validity, and temporal stability is lacking. The participation of experts in item generation is not detailed.
	Pilot	⊕	Construct validity	⊖	Pilot	⊕	
	Items	⊕	Convergent	⊖	Rigorous analysis	⊖	
Ensign et al. (2018)	Construct	⊕	Reliability	⊕	Steps	6/8	No CFA or test-retest No external expert review of item generation is specified.
	Pilot	⊕	Construct validity	⊖	Pilot	⊕	
	Items	⊕	Convergent	⊕	Rigorous analysis	⊕	
Garrido et al. (2018)	Construct	⊕	Reliability	⊕	Steps	6/8	Convergent validity and temporal stability (test-retest) were not evaluated. No mention of Cognitive pilot with target population.
	Pilot	?	Construct validity	⊕	Pilot	⊖	
	Items	⊕	Convergent	⊖	Rigorous analysis	⊕	
Morales et al. (2020)	Construct	⊕	Reliability	⊖	Steps	3/8	Information on reliability (Cronbach's alpha) and content validity is missing. Neither confirmatory factor analysis nor test-retest was performed. Item generation and pilot process not detailed.
	Pilot	?	Construct validity	⊖	Pilot	⊖	
	Items	⊖	Convergent	?	Rigorous analysis	⊖	
Ospina et al. (2022)	Construct	⊕	Reliability	?	Steps	3/8	The expert review process or Cognitive Pilot is not detailed in this study (although subsequent validations are cited). Neither temporal stability nor convergent validity is assessed.
	Pilot	?	Construct validity	⊖	Pilot	⊖	
	Items	⊕	Convergent	⊖	Rigorous analysis	⊖	

Note. ⊕ = The criterion is met; ⊖ = Criterion not met; ? = Insufficient information available.

Overall, the rigor of the procedures reported varied widely. Cognitive pretesting of items was conducted in only two instruments (Ensign et al., 2011; Ensign et al., 2018), while most studies did not provide evidence of this essential step. In terms of internal consistency, four instruments reported Cronbach's alpha values above .80 (e.g., Garrido-Hernansaiz et al., 2018, α = .94; Alonso et al., 2021, α = .81), whereas two studies did not report alpha at all (Barragán & Pérez-Jorge, 2020; Morales et al., 2020).

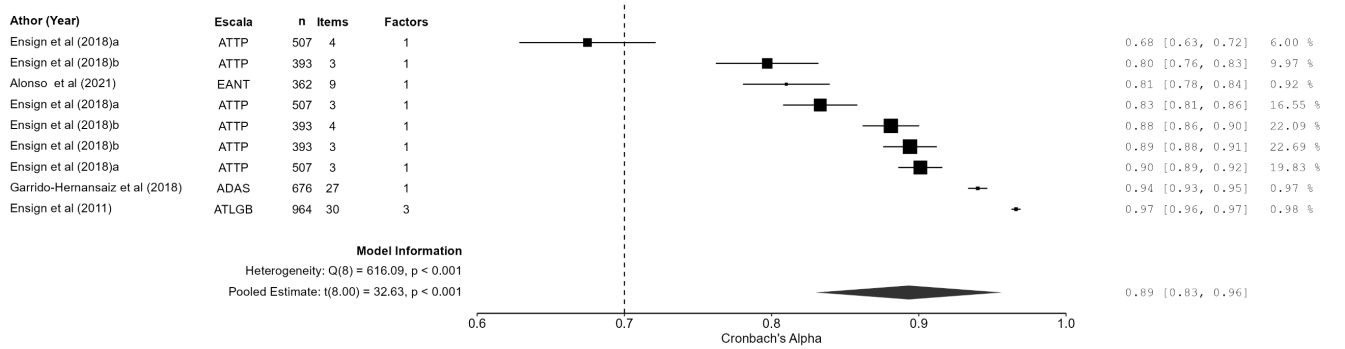
Regarding factorial validity, three instruments were evaluated with confirmatory factor analysis (CFA), and most of them showed acceptable fit indices (e.g., Alonso et al., 2021, CFI = .98). Others conducted only exploratory factor analysis (e.g., Ospina-Betancurt et al., 2023) or provided no structural evidence. Convergent validity was reported in just two studies, with moderate-to-strong correlations with related constructs (r = .72–.83; Ensign et al., 2018; Alonso et al., 2021). No study reported test–retest reliability, which limits evidence on the stability of scores over time. When mapped against DeVellis & Thorpe's recommended eight-step process, the strongest instruments (e.g., Garrido-Hernansaiz et al., 2018; Alonso et al., 2021) complied with approximately six of the steps, while the weakest (e.g., Morales et al., 2020; Barragán & Pérez-Jorge, 2020) satisfied only three or fewer.

In summary, while a few instruments reported acceptable factorial validity and internal consistency, none met the full set of international standards for scale development and validation. Across studies, item generation was often limited, Cronbach's alpha was used as the sole indicator of reliability, and no evidence of test-retest stability was provided. Taken together, these shortcomings suggest that the methodological quality of existing instruments is uneven and still falls short of what would be expected in contemporary psychometric research.

2.3.2. Reliability meta-analysis

In addition to the descriptive reporting of reliability coefficients, we conducted a quantitative meta-analysis of internal consistency to provide a reproducible, aggregated estimate across studies. This analysis was part of our a priori plan and was carried out following Bonett's transformation for reliability coefficients (ABT) under a multilevel random-effects model (see Appendix B for details).

Fig. 2 Forestplot of the reliability reported by the scales included in the review.



Across the seven studies, nine reliability estimates were extracted. The meta-analysis yielded an aggregated reliability of α = .90 (95% CI [.81, .95]), indicating high internal consistency overall. The standardized effect was ABT = 2.31, SE = 0.34, z = 6.86, p < .001, 95% CI [1.65, 2.98]. However, the results showed substantial heterogeneity (Q(8) = 1,060.15, p < .001; between-study variance σ^2= 0.57), suggesting marked variability across instruments.

As illustrated in the forest plot (Figure 2), almost all instruments achieved values above the conventional cut-off of .70, with most also exceeding the more stringent .80 threshold recommended for strong evidence of internal consistency (Nunnally & Bernstein, 1994; Tavakol & Dennick, 2011). Nevertheless, these results should be interpreted with caution. The small number of available effects and the incomplete reporting of psy-

chometric data in several studies limit the informativeness of the analysis. In practice, the meta-analysis adds little beyond the descriptive reporting provided in the previous subsection, though it demonstrates the feasibility of applying quantitative synthesis methods to this type of evidence.

In summary, the meta-analysis confirmed that most instruments reported reliability above accepted thresholds, with an aggregated estimate in the high range. At the same time, the strong heterogeneity and the very small evidence base mean that these findings should be read with caution. Although modest in this context, the inclusion of this analysis adds value by making the process transparent and offering a replicable framework that future reviews can build upon as more studies become available.

3. Discussion

The present study systematically reviewed instruments designed to assess attitudes toward sexual and gender diversity in educational contexts, guided by three research questions: (RQ1) which constructs these instruments measure and whether they capture both explicit and subtle attitudes; (RQ2) what descriptive characteristics define the studies in which they have been applied; and (RQ3) what methodological rigor and psychometric properties the instruments report. The findings highlight both contributions and limitations: most instruments focused primarily on explicit homophobic attitudes, with less attention to transphobia or broader diversity; the majority of studies were conducted with student samples in Spain and the United States; and although some tools reported strong internal consistency, none fully met international standards for scale development.

RQ1. Which constructs are targeted by the existing instruments, and how adequate are they for capturing both explicit and subtle attitudes?

The results showed that two instruments focused on attitudes toward sexual orientation, especially homosexuality and, to a lesser extent, bisexuality (ATLGB, ADAS). Only two of the reviewed instruments explicitly assessed attitudes toward gender identity, with an exclusive focus on transgender individuals (ATATP, EANT). The SASDA scale, used in the study by Ospina-Betancurt et al. (2023), evaluated attitudes toward both affective-sexual orientations and gender identities. Similarly, Alonso-Martínez et al. (2021) complemented their assessment of gender diversity with a measure of attitudes toward homosexuality to establish convergent validity, while Barragán-Medero and Pérez-Jorge (2020) also examined both dimensions of diversity. Notably, none of the reviewed instruments specifically addressed binegativity or microaggressions, despite their documented impact on LGBTIQ+ populations (Herek & McLemore, 2013; Sue, 2010).

Regarding the nature of attitudes, whether explicit or subtle, the ATLGB and ADAS scales rely primarily on overt prejudice and open rejection (e.g., "same-sex marriages should not be allowed"; Ensign et al., 2011; Garrido-Hernansaiz et al., 2018). Three instruments combine explicit and subtle forms of prejudice, such as stereotypes, indirect beliefs, or the absence of positive emotions, often balancing negatively worded items with affirmatively framed ones to mitigate social desirability bias (EANT, SASDA, and the MHS adaptation by Morales). Only two instruments can be clearly classified as measuring more subtle forms of attitudes. The ADAS and Barragán's ad hoc questionnaire incorporate positive dimensions such as respect, inclusion, and self-affirmation, which allow both the detection of attitudinal change and the identification of contexts that may benefit from educational or psychosocial interventions.

Although instruments targeting subtler attitudes generally rely on two strategies, balancing item valence and including indicators of modern or indirect prejudice, their application remains inconsistent, and many tools still rely on outdated or hostile statements. Notably, only two instruments can be clearly classified as measuring subtle forms of attitudes, underscoring the need for tools capable of capturing nuanced and socially mediated expressions of prejudice.

This limited coverage contrasts with instruments developed in other fields. For example, the Heterosexual Attitudes Toward Homosexuality Scale (HATH; Larsen et al., 1980), the Gender Role Conflict Scale (GRCS; O'Neil, 1986), and the Habarth Gender Role Belief Scale (Habarth, 2015) offer broader conceptualizations of gender-related bias and attitudes. Although these scales were not specifically designed for educational contexts, they illustrate how dimensions such as implicit bias, benevolence, and internalized conflict can be operationalized. The absence of equivalent tools in educational settings limits the ability to capture the full spectrum of attitudes that students and teachers may hold.

Beyond the instruments identified in school settings, additional tools developed in other contexts may also inform future measurement in education. For instance, Habarth's Gender Role Beliefs Scale (2015)

captures gender-related ideologies, while Massey's Sexual Prejudice Scale (2009) focuses on attitudes toward sexual minorities more broadly (Massey, 2009). Although not originally designed for educational populations, both instruments illustrate dimensions that could be adapted and validated for school contexts, thereby broadening the conceptual scope of future research.

RQ2. What are the descriptive characteristics of the included studies in which these instruments were applied?

Of the studies reviewed, five focused on the field of education (Ensign et al., 2011; Ensign et al., 2018; Morales et al., 2020; Ospina-Betancurt et al., 2023). Among these, Ensign et al. (2011) and Ensign et al. (2018) applied their instruments specifically within the context of university sports, whereas Ospina-Betancurt et al. (2023) examined attitudes among students of physical activity and sport sciences. Barragán-Medero and Pérez-Jorge (2020) administered their instrument to primary and secondary education students, and Garrido-Hernansaiz et al. (2018) focused exclusively on secondary education.

Overall, the reviewed studies primarily applied their instruments to student samples (Alonso-Martínez et al., 2021; Garrido-Hernansaiz et al., 2018; Morales et al., 2020; Ospina-Betancurt et al., 2023). Only one study included teachers in its sample (Ensign et al., 2011), while two others examined the attitudes of both teachers and students (Barragán-Medero & Pérez-Jorge, 2020; Ensign et al., 2018).

Geographically, most studies were conducted in Spain, with only two carried out in the United States, underscoring the limited cultural scope of the evidence. In addition, gender was reported almost exclusively in binary categories, with only isolated cases including a non-binary option, and sexual orientation was often reduced to a dichotomy of "heterosexual vs. non-heterosexual." Such practices limit inclusivity and ecological validity. Finally, most studies relied on convenience samples, with sample sizes ranging from fewer than 200 participants to more than 600. These methodological decisions restrict the generalizability of the findings to broader and more diverse populations.

RQ3. What is the methodological quality of the development process, and what psychometric properties have been reported according to international standards?

Among the reviewed studies, the psychometric rigor of the instruments varied considerably. Ensign et al. (2011) used Herek's widely validated scale (2000) but introduced modifications (reversing negatively worded items and adding a new dimension) without conducting additional reliability or validity analyses. By contrast, Ensign et al. (2018) developed a new instrument through a detailed and theoretically grounded validation process. Alonso-Martínez et al. (2021) reported the English linguistic validation of the EANT, including pilot testing and analyses of reliability, validity, and both exploratory and confirmatory factor structures. The SASDA (Piedra, 2016), used by Ospina-Betancurt et al. (2023), underwent validation through expert review, a pilot study, factor analyses, and reliability testing, with factors categorized according to established homohysteria levels. Garrido-Hernansaiz et al. (2018) detailed the construction and content validation of the ADAS, including item analyses, internal consistency, and factor analyses. Morales et al. (2020) reported that the MHS demonstrated satisfactory psychometric properties in educational contexts, although no specific analyses were presented. In contrast, Barragán-Medero and Pérez-Jorge (2020) provided no information on validity, factor structure, or reliability for their ad hoc self-assessment questionnaires, limiting their psychometric transparency.

Although some instruments reported high internal consistency and factorial validity, none fully adhered to international standards such as the COSMIN checklist or the APA Standards. Common deficiencies included the absence of cognitive pretests, limited item development, reliance on Cronbach's alpha as the sole indicator of reliability, and lack of test-retest evidence. The reliability meta-analysis confirmed that most instruments achieved coefficients above conventional cutoffs (.70 for acceptable, .80 for strong), but the high heterogeneity and limited number of effects imply that these findings should be interpreted with caution. Taken together, these results suggest that existing tools provide partial evidence of reliability and validity but fall short of the robustness expected in contemporary psychometric research.

4. Limitations of the study

This review also highlights the limitations of the available evidence. A narrow cultural scope, binary conceptualizations of gender, and the absence of instruments capturing subtle, modern forms of prejudice restrict the ecological validity of current measures. Future research

should prioritize the development of inclusive instruments that consider bisexual and non-binary identities, assess microaggressions, and incorporate multiple attitudinal dimensions (cognitive, affective, and behavioral). Moreover, new tools should be validated with diverse populations, including teachers, primary education students, and cross-cultural samples, and tested using rigorous psychometric procedures, such as confirmatory factor analyses with appropriate estimators, tests of measurement invariance, and test-retest reliability.

From an applied perspective, there is a pressing need for scales that are not only valid and reliable but also usable in teacher training and educational interventions, where they can guide practices to foster inclusive school climates. The development of such instruments would bridge the gap between academic research and practical application, ensuring that advances in measurement translate into meaningful changes in educational environments.

5. Conclusion

This systematic review provides a critical synthesis of instruments designed to measure attitudes toward sexual and gender diversity in educational contexts. Three main conclusions emerge.

First, the identified instruments capture explicit homophobic attitudes reasonably well but offer limited coverage of subtle, modern, or multidimensional forms of prejudice. Constructs such as binegativity, microaggressions, and non-binary or bisexual identities remain virtually absent, limiting the ecological validity of existing measures.

Second, the descriptive characteristics of the reviewed studies reveal a narrow evidence base. Most instruments have been applied to student populations in Spain and the United States, with little validation in other cultural contexts and very limited testing among teachers or younger students. In addition, demographic reporting often relied on binary categories, further restricting inclusivity.

Third, although some tools demonstrated acceptable levels of reliability and factorial validity, none fully met international standards for rigorous scale development and validation. Common methodological weaknesses included limited item development, reliance on Cronbach's alpha as the sole indicator of reliability, and the absence of test-retest data or evidence of measurement invariance.

Taken together, these findings highlight the need for future research to: (a) develop inclusive instruments that capture subtle and modern forms of bias; (b) expand validation efforts across diverse cultural and educational contexts; and (c) adopt rigorous psychometric procedures consistent with international frameworks (VALID, COSMIN, APA Standards, DeVellis & Thorpe, 2021). From an applied perspective, the development of robust tools that can be used in teacher training and educational interventions is essential to bridge the gap between academic research and practice and to promote more inclusive school environments.

Data Availability

All the material used throughout the review and analysis process is available through open and public access in the project created for this purpose in the OSF repository. (https://osf.io/qrtb5/)

CRediT authorship contribution statement

Conceptualization: D.S., M.A., and S.P.L.; Literature search: D.S.; Data coding: D.S., M.A., and S.P.L.; Data Curation, Software, Data analysis: S.P.L.; Writing - Review & Editing: D.S., M.A., and S.P.L.; Funding: D.S. and S.P.L., Supervision: M.A., and S.P.L.

Funding declaration

Funding declaration:
This work was supported by Junta de Andalucía under Grant HUM-642.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

We disclose that we used an AI-assisted tool for language polishing.

References

* (Marked with an * the studies included in the review).

Aggarwal, A., Qiao, S., O'Leary, S. D., Schlekat, K. N., & Li, X. (2024). Measurement Instruments Assessing Multi-Faceted Stigma Regarding Sexual and Gender Minorities: A Systematic Review of Psychometric Properties. *AIDS and behavior, 28*(6), 2054–2077. https://doi.org/10.1007/s10461-024-04305-2

Ajzen, I. (2001). *Nature and operation of attitudes*. *Annual Review of Psychology, 52*(1), 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Alegre, Y., & Fiedler, S. (2021). La categoría transgénero y su descontento: una genealogía crítica. *Revista Punto Género,* 16, 266–290. https://dx.doi.org/10.5354/2735-7473.2021.65895

*Alonso-Martínez, L., Heras-Sevilla, D., Fernández-Hawrylak, M., & Forrest, S. (2021). English Validation of a Short Scale Designed to Detect Negative Attitudes towards Trans People (EANT). *Sustainability, 13*(7), 3760. https://doi.org/10.3390/su13073760

American Psychological Association, & National Council on Measurement in Education (2020). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and communications board task force report. *American Psychologist, 73,* 3–25. https://psycnet.apa.org/doi/10.1037/amp0000389

*Barragán-Medero, F., & Pérez-Jorge, D. (2020). Combating homophobia, lesbophobia, biphobia and transphobia: A liberating and subversive educational alternative for desires. *Heliyon, 6*(10), e05225. https://doi.org/10.1016/j.heliyon.2020.e05225

Bartual-Figueras, M., Donoso-Vázquez, T., Sierra-Martínez, F., & Turmo, J. (2023). Validación de una escala de conciencia de género en el alumnado de ciencias de la salud. *Gaceta Sanitaria, 37*(0), 102304. https://doi.org/10.1016/j.gaceta.2023.102304

Bishop, C. J., & Pynoo, E. (2020). How well are outgroup attitudes and behaviours toward bisexual individuals measured? A systematic review of the psychometric properties of binegativity measures. *Psychology & Sexuality, 13*(2), 182–200. https://doi.org/10.1080/19419899.2020.1756390

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of educational and behavioral statistics, 27*(4), 335–340. https://doi.org/10.3102/10769986027004335

Bravestone, L., Hammond, M. D., Muise, A., & Cross, E. J. (2024). "It Wasn't Meant for Gays": Lesbian Women's and Gay Men's Reactions to the Ambivalent Sexism Inventory. *Sex Roles*, 90, 1585–1598. https://doi.org/10.1007/s11199-024-01528-4

Bulboacă, G., & Robu, V. (2021). A Psychometric Evaluation of the Revised Attitudes toward Lesbians and Gay Men Scale-Short Version: The ATLG-R-S-9. *Journal of Innovation in Psychology, Education and Didactics*, *25*(2), 225–238. https://doi.org/10.29081/JIPED.2021.25.2.10

Carrión, I., Bravo S., Izquierdo, S., & Marrero, E. (2020). Conocimientos sobre sexualidad y conductas sexuales de jóvenes universitarios, Cañar. *RECIMUNDO, 4*(4), 115–128. https://doi.org/10.26820/recimundo/4.(4).noviembre.2020.115-128

Compare, C., Rosati, F., Albanesi, C., Baiocco, R., & Lorusso, M. M. (2024). Embracing diversity: Exploring attitudes and beliefs toward transgender and gender-diverse people in the LGBTQIA + communities. *International Journal of Transgender Health*, First on line, 11–14. https://doi.org/10.1080/26895269.2024.2335517

Cramwinckel, F. M., Scheepers, D. T., & van der Toorn, J. (2018). Interventions to reduce blatant and subtle Sexual Orientation- and Gender Identity Prejudice (SOGIP): Current knowledge and future directions. *Social Issues and Policy Review*, *12*(1), 183–217. https://doi.org/10.1111/sipr.12044

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale Development: theory and applications*. SAGE.

Dovidio, J. F., Jones, J. M., & Vietze, D. L. (2025). *The psychology of diversity: Beyond prejudice and racism*. John Wiley & Sons.

*Ensign, K. A., Dodge, B. M., Herbenick, D., & Docherty, C. L. (2018). Development of an Instrument to Assess Athletic Trainers' Attitudes Toward Transgender Patients. *Journal of athletic training, 53*(4), 431–436. https://doi.org/10.4085/1062-6050-496-16

*Ensign, K. A., Yiamouyiannis, A., White, K. M., & Ridpath, B. D. (2011). Athletic trainers' attitudes toward lesbian, gay, and bisexual National Collegiate Athletic Association student-athletes. *Journal of Athletic Training, 46*(1), 69–75. https://doi.org/10.4085/1062-6050-46.1.69

Epps, B., Markowski, M., & Cleaver, K. (2023). A rapid review and narrative synthesis of the consequences of non-inclusive sex education in UK schools on lesbian, gay, bisexual, transgender and questioning young people. *Journal of School Nursing, 9*(1), 87–97. https://doi.org/10.1177/10598405211043394

European Commission (2023). *Progress report on the implementation of the LGBTIQ Equality Strategy 2020-2025.* Publications Office of the European Union. https://euro-cid.mne.gov.pt/sites/default/files/repository/content/event/45496/documents/justlgbtiq-strategyprogress-report-finalweb.pdf

European Union Agency for Fundamental Rights, FRA (2024). *LGBTIQ equality at a crossroads. Progress and challenges*. European Union Agency for Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2024-lgbtiq-equality_en.pdf

Formby, E. (2011). *Sexuality and gender in education: A critical introduction*. Routledge.

Garaigordobil, M., & Larrain, E. (2020). Bullying and cyberbullying in LGBT adolescents: Prevalence and effects on mental health. *Comunicar, 62*, 79–90. https://doi.org/10.3916/C62-2020-07

García-Orriols, J., & Torrebadella-Flix, X. (2019). Homophobia and sexual orientation in federated athletes in Catalonia: Immersed in the legacy of heteronormativity. *Masculinities and Social Change, 8*(3), 222–250. https://doi.org/10.17583/MCS.2019.4308

García-Pérez, R., Rebollo, M.A, Buzón, O., González-Piñal, R., Barragán, R., & Ruiz, E. (2010). Actitudes del alumnado hacia la igualdad de género. *Revista de Investigación, 28*(1), 217232. https://revistas.um.es/rie/article/view/98951

*Garrido-Hernansaiz, H., Martín-Fernández, M., Castaño-Torrijos, C., & Cuevas, I. (2018) Development and Validation of the ADAS Scale and Prediction of Attitudes Toward Affective-Sexual Diversity Among Spanish Secondary Students, *Journal of Homosexuality, 65*(8), 1032-1050. https://doi.org/10.1080/00918369.2017.1364951

Gromadzki, J. (2019). Measurement of attitudes towards homosexuality: testing measurement invariance across European countries. *Quality & Quantity, 53*(4), 2685-2694. https://doi.org/10.1007/s11135-019-00880-y

Habarth, J. M. (2015). Development of the Heteronormative Attitudes and Beliefs Scale. *Psychology & Sexuality, 6*(2), 166–188. https://doi.org/10.1080/19419899.2013.876444

Herek, G. M. (1998). The Attitudes Toward Lesbians and Gay Men (ATLG) scale. In C. M. Davis, W. L. Yarber, & R. Bauserman (Eds.), *Handbook of sexuality-related measures* (pp. 392–394). Sage Publications.

Herek, G. M. (2000). The psychology of sexual prejudice. *Current directions in psychological science*, *9*(1), 19–22. https://doi.org/10.1111/1467-8721.00051

Herek, G. M. (1988). Heterosexuals' attitudes toward lesbians and gay men: Correlates and gender differences. *The Journal of Sex Research*, **25**, 451–477. https://doi.org/10.1080/00224498809551476

Herek, G. M., & McLemore, K. A. (2013). Sexual prejudice. *Annual Review of Psychology, 64*, 309–333. https://doi.org/10.1146/annurev-psych-113011-143826

Hudson, W. W., & Ricketts, W. A. (1980). A strategy for the measurement of homophobia. *Journal of Homosexuality, 5*(4), 357–372. https://doi.org/10.1300/J082v05n04_02

Kerschbaumer, S., Voracek, M., Aczél, B., Anderson, S. F., Booth, B. M., Buchanan, E. M., ... & Tran, U. S. (2025). VALID: a checklist-based approach for improving validity in psychological research. *Advances in Methods and Practices in Psychological Science*, *8*(1), 25152459241306432. https://doi.org/10.1177/25152459241306432

Kosciw, J. G., Clark, C. M., & Menard, L. (2022). *The 2021 National School Climate Survey: The experiences of LGBTQ+ youth in our nation's schools*. GLSEN. Retrieved from (29/09/2025): https://eric.ed.gov/?id=ED625378

Lara-Garrido, A. S., Álvarez-Bernardo, G., & García-Berbén, A. B. (2024). Validity and measurement invariance of the Modern Homonegativity Scale for students in the field of education. *Annals of psychology, 40*(1), 85–94. https://doi.org/10.6018/analesps.544141

Larsen, K. S., Reed, M., & Hoffman, S. (1980). Attitudes of heterosexuals toward homosexuality: A Likert-type scale and construct validity. *Journal of Sex Research, 16*(3), 245–257. https://doi.org/10.1080/00224498009551081

López-Orozco, C. F., et al. (2022). A systematic mapping literature review of education around sexual and gender diversities. *Frontiers in Sociology, 7*, 946683. https://doi.org/10.3389/fsoc.2022.946683

Luu, M., Haardoerfer, R., Nehl, E. et al. (2025). Psychometric Properties of the Modern Homonegativity Scale among a Sample of Asian Americans. *Sexuality & Culture, 29*, 314–335. https://doi.org/10.1007/s12119-024-10272-6

Maschi, T., Malis, S., Christie, P., Adams, D., Goldstein, R., & Kaye, A. M. (2022). Gender identity and sexual orientation: Equally free to be LGBTQIA+: This is who I am, and we are! In C. Cox & T. Maschi (Eds.), *Human rights and social justice*, pp. 27–31. Routledge. https://doi.org/10.4324/9781003111269

Massey, S. G. (2009). Polymorphous prejudice: Liberating the measurement of heterosexuals' attitudes toward lesbians and gay men. *Journal of Homosexuality, 56*(2), 195–228. https://doi.org/10.1080/00918360802623131

Matsick, J. L., Palmer, L., Oswald, F., Kruk, M., & Ye, K. (2025). Illustrations of Benevolent and Hostile Heterosexism in LGBTQ+ People's Lives. *Journal of Homosexuality,* 1–27. https://doi.org/10.1080/00918369.2025.2480780

McKay, T., Lindquist, C. H., & Misra, S. (2017). Understanding (and Acting On) 20 Years of Research on Violence and LGBTQ + Communities. *Trauma, Violence, & Abuse*, *20*(5), 665–678. https://doi.org/10.1177/1524838017728708

Mella Gómez, L. E. (2024). El reconocimiento legal LGBTI y sus factores determinantes a nivel mundial: evidencia de datos de panel de 144 países. *Revista Española de Ciencia Política, 64*, 65–96. https://doi.org/10.21308/recp.64.03

*Morales Rodríguez, F. M., Rodríguez Clares, R., & García Muñoz, M. R. (2020). Influence of Resilience, Everyday Stress, Self-Efficacy, Self-Esteem, Emotional Intelligence, and Empathy on Attitudes toward Sexual and Gender Diversity Rights. *International journal of environmental research and public health, 17*(17), 6219. https://doi.org/10.3390/ijerph17176219

Morrison, M. A, & Morrison, T. G. (2003). Development and validation of a scale measuring modern prejudice toward gay men and lesbian women. *Journal of Homosexuality*, 43, 15–37. https://doi.org/10.1300/j082v43n02_02

Morrison, M. A., Bishop, C., & Morrison, T. G. (2018). A Systematic Review of the Psychometric Properties of Composite LGBT Prejudice and Discrimination Scales. *Journal of Homosexuality*, *66*(4), 549–570. https://doi.org/10.1080/00918369.2017.1422935

Muñoz-García, L. E., Gómez-Berrocal, C., Guillén-Riquelme, A., & Sierra, J. C. (2023). Measurement Invariance across Sexual Orientation for Measures of Sexual Attitudes. *International journal of environmental research and public health, 20*(3), 1820. https://doi.org/10.3390/ijerph20031820

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

O'Neil, J. M., Helms, B. J., Gable, R. K., David, L., & Wrightsman, L. S. (1986). Gender Role Conflict Scale: College men's fear of femininity. *Sex Roles, 14*(5–6), 335–350. https://doi.org/10.1007/BF00287583

OECD (2023). *Fighting homophobia and transphobia in schools: A groundbreaking impact assessment*. OECD Publishing, Paris. https://doi.org/10.1787/4950506c-en.

*Ospina-Betancurt, J., Martínez-Patiño, M., Tovar Cuevas, J., & García-Rubio, J. (2023). Attitudes toward sexual diversity in sport among undergraduate students of physical activity and sport sciences in Spain. *Sexuality & Culture, 27*, 878–893. https://doi.org/10.1007/s12119-022-10043-1

Özdere, M. (2023). School Violence Against LGBT-Q: A Review Study. *ISPEC International Journal of Social Sciences & Humanities, 7*(2), 363–381. https://doi.org/10.5281/zenodo.7987868

Páez, J., Hevia, G., Pesci, F., & Rabbia, H. H. (2015). Construcción y validación de una escala de actitudes negativas hacia personas trans. *Revista de Psicología, 33*(1), 153–190. https://doi.org/10.18800/psico.201501.006

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, *372*. https://doi.org/10.1136/bmj.n71

Payne, E. C., & Smith, M. J. (2018). Refusing relevance: School administrator resistance to offering professional development addressing LGBTQ issues in schools. *SAGE Open, 8*(4), 215824401880456. https://doi.org/10.1177/2158244018804564

Piedra, J. (2016). Attitude scale towards sexual diversity in sport (EDSD): Development and preliminary validation. *Revista de Psicología del Deporte, 25*(2), 299–307. https://doi.org/10.5232/ricyde2016.02502.06

Shoaei, K. (2021). *Discrimination against LGBTQ students in United States higher education: A qualitative case study* (Doctoral dissertation). Northcentral University. ProQuest Dissertations & Theses Global.

Stanick, C., Halko, H., Mettert, K., Dorsey, C., Moullin, J., Weiner, B., Powell, B., & Lewis, C. C. (2021). Measuring characteristics of individuals: An updated systematic review of instruments' psychometric properties. *Implementation Research and Practice, 2*. https://doi.org/10.1177/26334895211000458

Stiekema, E., Warrens, M. J., van Bergen, D. D., & Parlevliet, S. (2024). Assessing attitudes toward LGB people in young adolescents. *PLOS ONE, 19*(10), e0305057. https://doi.org/10.1371/journal.pone.0305057

Suarez, N. A., Trujillo, L., McKinnon, J., II, Barrios, L. C., DePedro, K. T., Rasberry, C. N., Roberts, G., & Underwood, J. M. (2024). Disparities in school connectedness, unstable housing, experiences of violence, mental health, and suicidal thoughts and behaviors among transgender and cisgender high school students — Youth Risk Behavior Survey, United States, 2023. *MMWR Supplements, 73*(Suppl-4), 50–58. https://doi.org/10.15585/mmwr.su7304a6

Sue, D. W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation*. Wiley

Swan, K. (2023). Measuring what matters in healthcare: A practical guide to psychometric principles and instrument development. *Frontiers in Psychology, 14*, 1225850. https://doi.org/10.3389/fpsyg.2023.1225850

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2,* 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., ... & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research*, *27*, 1159-1170. https://doi.org/10.1007/s11136-018-1829-0

UNESCO (2020). *Global Education Monitoring Report 2020: Inclusion and education - All means all*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000373718

United Nations Educational, Scientific and Cultural Organization. (2019). *Bringing it out in the open: Monitoring school violence based on sexual orientation, gender identity or gender expression in national and international surveys* (Technical brief). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000367493

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software*, *36*, 1-48. https://doi.org/10.18637/jss.v036.i03

Wang, J., Liu, Y., Fu, G., Chen, Y., Wu, L., Pan, M., Yang, Y., Chen, Z., Cao, Y., Li, Y., Wang, H., Wang, B., Du, R., Xiong, Y., Liu, W., Xu, N., Xia, X., Li, Q., Lv, C., & Ruan, F. (2022). Validation of Herek's attitudes toward lesbian women and gay men scale among undergraduates in mainland China. *Frontiers in Psychology, 13,* Article 842397. https://doi.org/10.3389/fpsyg.2022.842397