

REVIEW ARTICLE

Psychometric Properties of Instruments Assessing Ageism: A Systematic Review

Chien-Chih Liu¹ | Yi-Jung Chen² | Xavier C. C. Fung³ | Iqbal Pramukti⁴ | Chieh-hsiu Liu⁵ | Li-Fan Liu⁶ | Mark D. Griffiths⁷ | Yi-Ching Yang^{5,8,9} | Chung-Ying Lin^{2,4,10,11} 

¹Center for General Education, National Tainan Junior College of Nursing, Tainan, Taiwan | ²Institute of Allied Health Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan | ³Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Hong Kong | ⁴Department of Community Health Nursing, Faculty of Nursing, Universitas Padjadjaran, Sumedang, West Java, Indonesia | ⁵Department of Geriatrics and Gerontology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan | ⁶Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan, Taiwan | ⁷International Gaming Research Unit, Psychology Department, Nottingham Trent University, Nottingham, UK | ⁸Department of Family Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan | ⁹Department of Family Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan | ¹⁰School of Nursing, College of Nursing, Kaohsiung Medical University, Kaohsiung, Taiwan | ¹¹Biostatistics Consulting Center, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Correspondence: Li-Fan Liu (lilian@mail.ncku.edu.tw) | Chung-Ying Lin (cylin36933@gmail.com)

Received: 29 April 2024 | **Revised:** 26 January 2025 | **Accepted:** 9 March 2025

Funding: This research was supported in part by (received funding from) the National Health Research Institutes (NHRI-11A1-CG-CO-04-2225-1) and the Academia Sinica (AS-GCS-113-H01). The funding body had no role in the present study.

Keywords: ageism | COSMIN | older people | psychometric instruments | systematic review

ABSTRACT

Introduction: The global population is experiencing a significant increase in the number of older people, highlighting the need to maintain both physical and mental health among this cohort and to promote healthy ageing. One critical area that has been insufficiently explored is the prevalence and scope of ageism and its assessment. Therefore, the present review evaluated the psychometric properties of instruments designed to assess ageism against older people.

Methods: To locate relevant instruments, a search was conducted using seven databases comprising *MEDLINE*, *PubMed*, *Embase*, *PsycINFO*, *Web of Science*, *ScienceDirect* and *Scopus*. The review utilised the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist to assess the methodological quality of the studies identified, ensuring a thorough evaluation of relevant literature.

Results: From an initial pool of 338,180 outputs, 20 studies were deemed suitable for inclusion. These studies addressed various aspects of ageism, with eight focusing on self-directed ageism and 12 on attitudes towards older individuals. The review found that while 18 studies tested internal consistency, only three assessed cross-cultural validity. None of the studies achieved a high methodological quality rating concerning content validity based on the COSMIN criteria.

Conclusions: These findings suggest a significant gap in the development of reliable and comprehensive psychometric instruments for assessing ageism, indicating a need for further refinement of these instruments. **Implications for Practice:** This systematic review equips healthcare policymakers with evidence to advance the development of more robust ageism measurement

Chien-Chih Liu and Yi-Jung Chen made equal contribution to be the co-first authors of the present paper.

tools. Additionally, the results underscore the need for researchers to conduct more thorough validation and evaluation of ageism instruments that address its multidimensional nature, thereby enhancing the quality of future gerontological research.

Implications for Practice: This systematic review equips healthcare policymakers with evidence to advance the development of more robust ageism measurement tools. Additionally, the results underscore the need for researchers to conduct more thorough validation and evaluation of ageism instruments that address its multidimensional nature, thereby enhancing the quality of future gerontological research.

1 | Introduction

The global population is rapidly ageing (United Nations 2017). Therefore, the maintenance of physical and mental health (i.e. healthy ageing and intrinsic capacity maintenance) (Chen et al. 2023) among older people has become an important issue. One of the most important topics associated with healthy ageing and intrinsic capacity is ageism. Ageism is regarded as a public health issue, and ageist attitudes or beliefs toward older adults among different parties (including the society, healthcare workers and older adults themselves) have been shown to have an association with their physical and mental health outcomes (Ayalon et al. 2019). Butler (1969) introduced the concept of ageism, which referred to one age group being prejudiced by other age groups. Cherry and Palmore (2008) described ageism as a type of prejudice or discrimination among society or human beings in favour of or against a specific age group. Moreover, they indicated that prejudices were a reflection of emotions derived from stereotypes or attitudes towards a specific age group, while discrimination as subsequently inappropriate behaviour towards people in a specific age group was due to negative prejudices, stereotypes or attitudes.

The World Health Organization (WHO 2021) has also noted that ageism can be self-directed, interpersonal and institutional. Self-directed ageism happens when ageism is internalised and individuals turn against themselves. Interpersonal ageism emerges from interactions between two or more individuals. Institutional ageism refers to laws, policies and practices of institutions that unfairly treat individuals based on their age.

Ageist attitudes may deleteriously affect large populations of older adults around the world. Older people who are exposed to negative age-related stereotypes can show a decline in functional performance (Smith et al. 2017). Moreover, Lee et al. (2012) have shown that stereotypes against older adults can lead to older people being disadvantaged and improperly treated in the health system. Ageism in healthcare persists everywhere from social interactions and organisational cultures to health policies (de São José and Amado 2017). Studies have also shown that older people can be unpopular among healthcare staff because of societal ageist attitudes (Ben-Harush et al. 2017). Gebremariam and Sadana (2019) indicated that older people often suffer indignities and unequal treatment from their healthcare providers.

To the best of the present authors' knowledge, there are few studies emphasising a systematic review of instruments that assess the prevalence and extent of ageism and their psychometric properties. Without reliable and valid instruments to assess ageism, healthcare providers and relevant stakeholders cannot identify the problems associated with ageism and are therefore unable to take actions to eliminate it. Therefore, it is necessary to review the

psychometric evidence regarding ageism assessment instruments, which can help address how ageism is expressed in the healthcare system and how it is experienced by older people in healthcare.

In order to address this literature gap, the present review aimed to identify the scope and psychometric properties of ageism assessment instruments. The results could provide critical information to identify the comprehensiveness of psychometrically valid scales that explore how prevalent ageism is in healthcare settings and society. Moreover, this information might be of great importance in reducing ageism. This information could inform evidence-based policies, training programmes, and public awareness campaigns, ultimately fostering a more inclusive and equitable environment for older individuals and helping to reduce ageism at both individual and systemic levels.

2 | Methods

A systematic review was performed based on the guidelines of both the (i) Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement for systematic reviews (Moher et al. 2009), and the (ii) CONsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist (Terwee et al. 2007). The systematic review protocol was registered on the international prospective register of systematic reviews (OSF Registries, registration doi: [10.17605/OSF.IO/KAUH6](https://doi.org/10.17605/OSF.IO/KAUH6)).

2.1 | Search Strategy

To conduct the literature review, *MEDLINE*, *PubMed*, *Embase*, *PsycINFO*, *Web of Science*, *ScienceDirect* and *Scopus* were utilised as they are considered the most relevant databases pertinent to the field of geriatrics, healthcare, and psychology. Ageism has been defined as 'stereotypes, prejudice, or discrimination against individuals or groups based on their age' (Butler 1969), and this definition guided the selection of studies in the present review. Specific aspects of ageism, such as self-directed ageism (i.e. older people devaluing themselves), interpersonal ageism (i.e. people interacting with older people being unfriendly), and institutional ageism (i.e., regulations or laws that are unfriendly to [or discriminate against] people of older age), were also considered to ensure a comprehensive review of the construct. This definition was incorporated into the review process to ensure consistency in identifying studies and interpreting their findings. The reviewed studies included those that explicitly addressed one or more aspects of ageism and used validated psychometric instruments to assess them. To search for relevant studies, a search strategy was

Summary

- What does this research add to existing knowledge in gerontology?
 - The present study highlights the lack of robust psychometric instruments for assessing ageism.
 - The findings indicate a significant gap in understanding how ageism is quantitatively evaluated in gerontological research.
 - The findings emphasise the need for comprehensive methodologies in studying ageism.
- What are the implications of this new knowledge for nursing care for and with older adults?
 - Nurses and healthcare professionals can better identify ageist practices and attitudes with improved assessment instruments.
 - Enhanced awareness and assessment of ageism can lead to more empathetic, respectful care for older people.
 - Training programmes can be developed to integrate findings to cultivate anti-ageist practices in nursing curriculums.
- How could the findings be used to influence practice, education, research, and policy?
 - Policymakers in healthcare settings can use the findings as a basis for promoting the development of more effective ageism assessment instruments.
 - Researchers are encouraged to validate and evaluate instruments on ageism more comprehensively (i.e. addressing different aspects of ageism) to improve future gerontological studies.

implemented based on the COSMIN approach. The approach involved constructing search filters for the key elements of the construct of interest: population(s), type of instruments (e.g., scale or questionnaire), and measurement properties (including inclusion and exclusion filters), and then combining them using AND and NOT Boolean operators. The search filter for population(s) was not applied because the study aimed to review all self-reported ageism instruments without considering specific populations. For the type of instruments and measurement properties, a modified filter developed by the Patient-Reported Outcomes Measurement Group at the University of Oxford and a validated highly sensitive search filter developed using the COSMIN were used (Terwee et al. 2009). The search was conducted up until 5 April 2024, utilising the following search strategy: Ageism [All fields] OR agism [All fields] OR ageist [All fields] OR agist [All fields] OR age discrimination [All fields] OR age prejudice [All fields] OR age stereotype [All fields] OR self-perceptions of aging [All fields] OR age identity [All fields] (see Appendix S1). All included studies were reviewed manually to identify any additional studies that may have been overlooked during the database search.

2.2 | Study Selection

The present review included all study designs, provided that the primary peer-reviewed paper reported at least one psychometric property associated with an instrument that assesses

ageism. These properties were defined by the COSMIN checklist and included structural validity, cross-cultural validity, internal consistency, criterion validity, measurement error, reliability, hypothesis testing for construct validity, content validity, and responsiveness. The review was confined to studies published in the English language and excluded grey literature such as unpublished literature, research theses or conference proceedings due to the potential lack of peer review and inadequate information to evaluate methodological quality. The results of all searches were entered into the EndNote Team, 2013, Philadelphia, U.S., Clarivate, Endnote X20.0.1 software program for systematic reviews. The titles and abstracts of potential studies were screened by two experienced psychometric research practitioners, C-CL and Y-JC. Any problems were resolved through discussion between the two reviewers. C-YL was consulted as a third reviewer if required. All potential papers were subjected to full-text screening using the procedure above. The validity of the findings was subsequently verified by another independent investigator, C-YL, for the final review.

2.3 | Evaluation of Psychometric Properties

Using the COSMIN Risk of Bias checklist (<http://www.cosmin.nl>), the present review assessed included studies' methodological quality. Unlike other risk of bias assessment instruments that focus on only a few aspects, the COSMIN checklist provides a comprehensive assessment. Therefore, it has been previously utilised in systematic reviews to evaluate the methodological quality of studies that involve performance-based outcomes with generic items designed for multiple applications. The COSMIN checklist comprises 10 aspects (internal consistency, reliability, criterion validity, measurement error, hypotheses testing [construct validity], structural validity, content validity, cross-cultural validity, patient-reported outcome measures development and responsiveness) that assess the methodological standards of a given study. A four-point rating scale (i.e. 'very good', 'adequate', 'doubtful' and 'inadequate') is employed to score on each item. In order to gauge the overall quality of a study, the principle of 'worst score counts' was employed. This principle asserts that the lowest rating of any standard within the set is taken into consideration. If a study is rated as 'inadequate' for any item, its overall quality is deemed inadequate. The COSMIN checklist also includes statistical quality evaluation criteria for different psychometric properties, as outlined in Appendix S2. Moreover, the study results corresponding to each criterion were rated as indeterminate (?), positive (+) or negative (−). As an illustration, a positive (+) rating is given if ICC ≥ 0.70 for reliability; a negative (−) rating if ICC ≤ 0.70 ; an indeterminate (?) rating if no reported ICC.

2.4 | Assessments of the Quality of Statistical Findings

An assessment was conducted on the statistical outcomes of each measurement property from the instruments on ageism in comparison with the updated criteria for adequate measurement properties specified by Prinsen et al. (2016) and Terwee

et al. (2012). The results were classified as satisfactory (+), inadequate (−) or inconclusive (?), and all evaluations regarding their corresponding criteria are presented in Appendix S2.

2.5 | Synthesis of Findings

The present systematic review did not include a meta-analysis. The synthesis procedure was conducted as follows. First, it used the COSMIN checklist to evaluate the quality of each included study. Second, it used the criteria for the evaluation of psychometric properties proposed by both Prinsen et al. (2016) and Terwee et al. (2012) to evaluate each psychometric property for each ageism measure. Third, it used the findings obtained from the aforementioned first and second steps to synthesise the psychometric properties of each ageism measure.

3 | Results

3.1 | Study Selection

A total of 338,180 records were retrieved from various databases and registers. After removing 132,724 duplicate entries, 205,456 records remained for screening. Among these, 30,818 were excluded because they did not mention any psychometric properties information in the title or abstract.

Of the 174,638 reports identified for eligibility assessment, 11 could not be retrieved because they were contained in conference proceedings or theses. Consequently, 174,627 records were evaluated in detail for inclusion. Of these, 104,414 were excluded for not focusing on psychometric properties (i.e. they may have reported some psychometric properties but the primary focus of study was not psychometric evaluation), and 70,192 were excluded because they were review papers (overviewing ageism and/or describing the effectiveness of ageism reduction programmes). Only two review papers have previously reviewed ageism instruments (i.e. Ayalon et al. 2019; Hu et al. 2021). The exclusion of review papers ensured the analysis was based exclusively on primary data, thereby minimising biases or interpretations that might arise from secondary sources. This approach enabled a direct evaluation of psychometric properties—such as validity, reliability, and responsiveness—as reported in original studies.

Review papers were also excluded to avoid potential issues of overlapping study summaries, which could inflate the evidence base or lead to double counting of data. This methodological decision aligns with COSMIN guidelines, which prioritise evidence derived from original studies for the evaluation of measurement properties. Following these criteria, a total of 20 studies were ultimately included in the review (Figure 1). Eight of the instruments are used to assess self-directed ageism (Allen et al. 2022; Barker et al. 2007; Gendron et al. 2020; Kang and Chasteen 2009; Laidlaw et al. 2007; Lasher and Faulkender 1993; Lawton 1975; Sarkisian et al. 2005), whereas the remaining 12 are used to assess attitudes toward old people. Table 1 summarises the samples' sociodemographics and characteristics for the included studies, whereas Tables 2–4 report the COSMIN grading and statistical findings. Table 5 presents the levels of

evidence for different measurement properties of ageism in the various studies.

3.2 | Content Validity

Of the 16 included papers that addressed content validity (Allen et al. 2022; Barker et al. 2007; Braithwaite et al. 1993; Cary et al. 2017; Gendron et al. 2020; Gething 1994; Kafer et al. 1980; Kang and Chasteen 2009; Kogan 1961; Laidlaw et al. 2007; Lasher and Faulkender 1993; Levy et al. 2004; Marchetti et al. 2022; North and Fiske 2013; Sarkisian et al. 2005; Tuckman and Lorge 1958), six used qualitative methods (Barker et al. 2007; Kang and Chasteen 2009; Laidlaw et al. 2007; Levy et al. 2004; Marchetti et al. 2022; Sarkisian et al. 2005). Item generation and selection procedures often included a combination of literature review, development of an overinclusive initial item pool, review of items for clarity, examination of item factor loadings and elimination of items with low factor loadings. Six scales had a development study, and one had a Delphi exercise.

Although almost all scale development studies described the construct, the target population and the context of use, only five conducted a focus group (Barker et al. 2007; Kang and Chasteen 2009; Laidlaw et al. 2007; Marchetti et al. 2022; Sarkisian et al. 2005), and none of these adequately asked the target population or professionals about the comprehensibility or comprehensiveness of the scale. Therefore, none of the scale development studies were of good quality. Sufficient evidence of content validity was found in only one study (Laidlaw et al. 2007). However, the quality of this evidence was low because it was unclear whether all items were tested with participants in their final form, and the target population was not asked about the comprehensiveness of the items.

3.3 | Structural Validity

Of the 20 studies, seven of them demonstrated very good methodological quality and had a negative rating for the quality of statistical findings based on a reflective model (Allen et al. 2022; Barker et al. 2007; Cary et al. 2017; Gendron et al. 2020; Laidlaw et al. 2007; Lasher and Faulkender 1993; Sarkisian et al. 2005) (Table 2). Five studies' results could not be scored on the unidimensionality criteria (Axelrod and Eisdorfer 1961; Braithwaite et al. 1993; Levy et al. 2004; Palmore 1980; Rosencranz and McNevin 1969). This was because of incomplete reporting of EFA eigenvalues or explained variance. Studies of at least adequate quality for PGC Morale Scale (Lawton 1975), RSQ-Age (Kang and Chasteen 2009) and Tuckman–Lorge Attitudes Toward Older People (Axelrod and Eisdorfer 1961) showed indeterminate results, indicating uncertain unidimensionality underpinned by low-quality evidence (Table 2).

3.4 | Internal Consistency

Eighteen studies evaluated the internal consistency and had very good methodological quality (Allen et al. 2022; Barker et al. 2007; Braithwaite et al. 1993; Cary et al. 2017; Fraboni et al. 1990; Gendron et al. 2020; Gething 1994; Kafer et al. 1980; Kang and

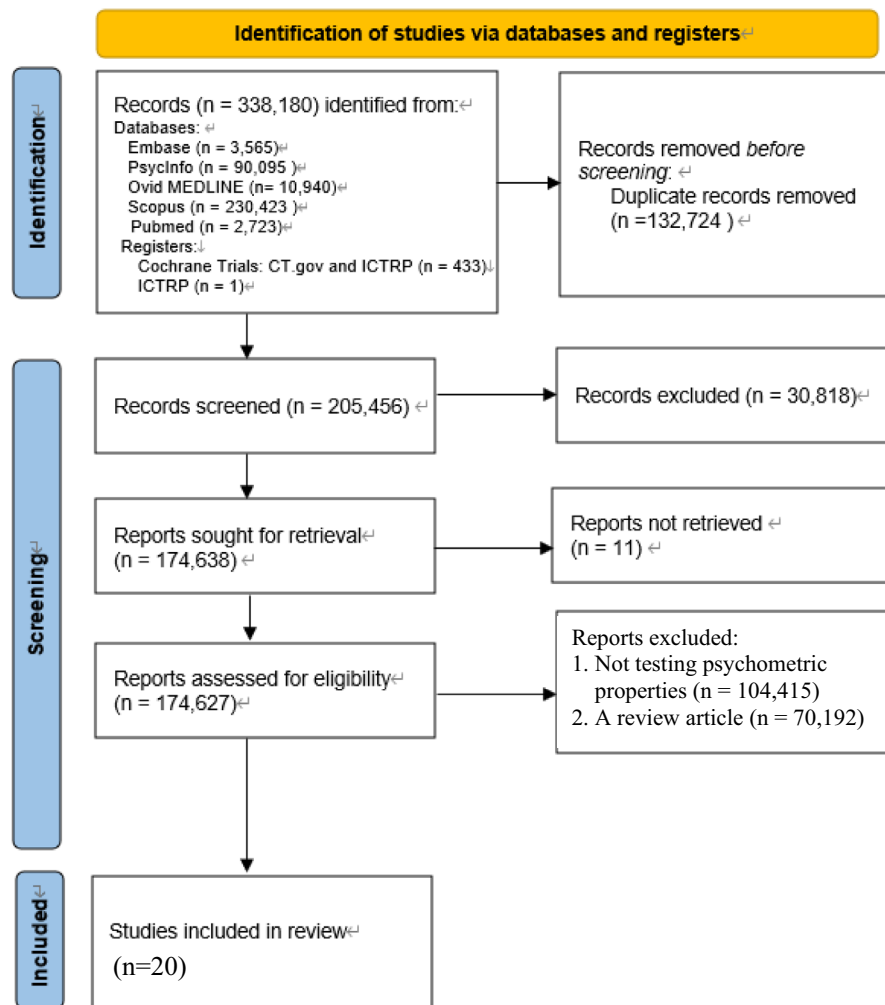


FIGURE 1 | PRISMA 2020 flow diagram of the study selection process. The main reason for exclusion was that the primary focus was not on psychometric properties. Moreover, 11 studies could not be retrieved because they were contained in conference proceedings or theses.

Chasteen 2009; Kogan 1961; Laidlaw et al. 2007; Lasher and Faulkender 1993; Lawton 1975; Levy et al. 2004; Marchetti et al. 2022; North and Fiske 2013; Palmore 1980; Sarkisian et al. 2005). Internal consistency was determined using Cronbach α calculations for these studies, the scores of which ranged from 0.47 for the Facts on Aging Quiz (Palmore 1980) to 0.91 for the RSQ-Age (Kang and Chasteen 2009). Studies describing the Prescriptive Intergenerational-Tension Ageism Scale (North and Fiske 2013) and the Ambivalent Ageism Scale (Cary et al. 2017) showed a positive rating for the quality of statistical findings indicating that the total score predominantly reflected a single factor. Five studies demonstrated negative quality criteria of the Pearson correlation coefficient attributed to the multidimensionality caused by the sub-domain factors (Barker et al. 2007; Braithwaite et al. 1993; Kafer et al. 1980; Kogan 1961; Palmore 1980). The best-evidence synthesis showed limited positive evidence for the internal consistency of the ageism instruments, using the same items (see Table 5).

3.5 | Cross-Cultural Validity and Measurement Invariance

Three studies (Kogan 1961; Laidlaw et al. 2007; North and Fiske 2013) examined the cross-cultural validity of the ageism

instruments across different factors (Kogan 1961; Laidlaw et al. 2007; North and Fiske 2013) (Table 3). The Attitudes to Ageing Questionnaire (Laidlaw et al. 2007) examined the effects of region, gender and age. To some extent, each measure compared scores with suitable and related variables. Kogan's Attitudes Towards Old People Scale (Kogan 1961) examined how different universities in the USA responded to these issues. All the other cross-cultural validity measures displayed negative ratings (Table 3).

3.6 | Reliability

Interrater reliability was conducted in one study (Laidlaw et al. 2007; ICC = 0.10), and seven studies conducted test-retest reliability (Axelrod and Eisdorfer 1961; Barker et al. 2007; Cary et al. 2017; Kang and Chasteen 2009; Laidlaw et al. 2007; Levy et al. 2004; Marchetti et al. 2022; ICC = 0.16–0.76). Among these, five studies reported test-retest reliability coefficients > 0.70 (Barker et al. 2007; Cary et al. 2017; Kang and Chasteen 2009; Levy et al. 2004; Marchetti et al. 2022), and one study reported an interrater reliability coefficient < 0.70 (Laidlaw et al. 2007). Notably, one study employed Rasch analysis to assess item properties (Laidlaw et al. 2007), finding that comparisons between European and non-European

TABLE 1 | The characteristics and sociodemographic of the populations in the included ageing instrument studies.

	Author(s)/ country	Study design	Sample size (Percentage of females)	Age (years)	Number of items	The COSMIN measurements addressed	Others (not included in the COSMIN checklist)
1	Philadelphia Geriatric Center (PGC) Morale Scale Lawton (1975) USA	Cross-sectional	828 (72% female)	72.6	17-item	- Structural validity - Internal consistency	NA
2	Anxiety about Aging Scale (AAS) Lasher and Faulkender (1993) USA	Cross-sectional	312 (61.22% female)	18–80	20-item	- PROM design - Content validity - Internal consistency - Responsiveness	Two-way analysis of variance (sex $\times$ age) on the total AAS produced a significant main effect for sex ($F[1,289] = 4.30$, $p < .05$), indicating that males scored higher on the AAS than females
3	Expectations Regarding Aging-12 Item (ERA-12) Sarkisian et al. (2005) USA	Cross-sectional	429 (54% female) 643 (77% female)	76.6 $\pm$ 6.9 77.5 $\pm$ 6.5	12-item	- PROM design - Content validity - Structural validity - Convergent validity	NA
4	Ageing Perception Questionnaire (APQ) Barker et al. (2007) Ireland	Cross-sectional	2033 (57% female)	74.1 $\pm$ 6.8	32-item	- Content validity - Structural validity - Reliability - Measurement error - Construct validity	Ceiling effect: 37.7% of participants who had experienced health-related changes attributed all of these health-related changes to ageing. Subsequent testing for the fourth condition of mediation showed that the inclusion of APQ dimensions yielded significant reductions in the beta-coefficients for comorbidity with Sobel's tests supporting a mediating role for individual APQ variables (timeline chronic, $z = 9.53$, $p < 0.00001$; emotional representations, $z = 9.91$, $p < 0.00001$; control positive, $z = 8.72$, $p < 0.00001$; consequences negative, $z = 11.35$, $p < 0.00001$)

(Continues)

TABLE 1 | (Continued)

	Author(s)/ country	Study design	Sample size (Percentage of females)	Age (years)	Number of items	The COSMIN measurements addressed	Others (not included in the COSMIN checklist)
5	<i>Attitudes to Aging Questionnaire (AAQ)</i>						
	Laidlaw et al. (2007)	Cross-national	1356 (60% female)	60–100	24-item	– PROM design – Content validity – Structural validity – Internal consistency – Cross-cultural validity – Convergent validity	Low rates of missing values across the items (range: 1.4%–4.5%). Values for skew and kurtosis showed few items displaying extreme distribution problems, though Items 12, 18, 22 and Item 28 showed some response distribution problems (see Table 5).
6	<i>Age-based Rejection Sensitivity Questionnaire (RSQ-Age)</i>						
	Kang and Chasteen (2009)	Pilot study	More than 2300 adults (70% female)	71.37 ± 9.89 (29– 102 years)	15-item	– PROM design – Content validity – Internal consistency – Reliability – Convergent validity	NA
7	<i>Relational Aging Anxiety Scale (RAA)</i>						
	Gendron et al. (2020)	Cross-sectional	329 (86% female identified)	49.2 ± 14.6 (21– 79 years)	16-item	– PROM design – Content validity – Structural validity – Internal consistency – Measurement invariance	The relationship between age, gender and negative impact of ageism were all significant; age chi-squared (15, $N = 320$) = 31.93, $p < 0.007$, gender chi-squared (15, $N = 327$) = 22.61, $p < 0.007$. Collective affinity for older people ($\beta = 0.22$, $p < 0.001$) and relational ageing anxiety ($\beta = 0.21$, $p = 0.001$) each significantly predicted knowledge of negative health outcomes on ageism
8	<i>Everyday Ageism Scale</i>						
	Allen et al. (2022)	Cross-sectional	2048 (52.4% female)	50–80	10-item	– Structural validity – Internal consistency – Convergent validity – Known-groups validity – Construct validity	

(Continues)

TABLE 1 | (Continued)

	Author(s)/ country	Study design	Sample size (Percentage of females)	Age (years)	Number of items	The COSMIN measurements addressed	Others (not included in the COSMIN checklist)
9	<i>Tuckman-Lorge Attitudes Toward Old People</i> Axelrod and Eisdorfer (1961) USA	Cross-sectional	28.0 (38.21% female)	18–22	137-item	– Content validity	Phi coefficient derived from the pooled percentages
10	<i>Kogan Attitudes Toward Old People Scale</i> Kogan (1961) USA	Cross-sectional	482 (16.8% female)	18–22	17 matched positive–negative pairs items	– Content validity – Internal consistency – Convergent validity	NA
11	<i>Aging Sematic Differential</i> Rosencranz and McNevin (1969)/ USA	Cross-sectional	287 (specified males)	18–22	32 matched positive–negative pairs items	– Content validity – Measurement invariance	NA
12	<i>Aging Opinion Survey</i> Kafer et al. (1980) USA	Cross-sectional	440 (majority female)	18–22	15-item	– Content validity – Structural validity – Internal consistency – Measurement invariance	NA
13	<i>Facts on Aging Quiz</i> Palmore (1980) USA		3351	18–22	25-item	– Internal consistency – Reliability	Face validity
14	<i>Fraboni Scale of Ageism</i> Fraboni et al. (1990) Canada	Cross-sectional	231 (65.37% female)	18–22	29-item	– Structural validity – Internal consistency – Convergent validity	NA
15	<i>Awareness of ageism</i> Braithwaite et al. (1993) Australia	Cross-sectional	195 (70% female)	18–22	10-item	– Content validity – Structural validity – Internal consistency – Measurement invariance – Hypotheses testing	NA

(Continues)

TABLE 1 | (Continued)

	Author(s)/ country	Study design	Sample size (Percentage of females)	Age (years)	Number of items	The COSMIN measurements addressed	Others (not included in the COSMIN checklist)
16	<i>Reactions to Ageing Questionnaire (RAQ)</i>						
	Gething (1994) Australia	Cross-sectional	531 (88.4% female)	30–49	30-item	– Structural validity – Internal consistency – Measurement invariance – Convergent validity	NA
17	<i>Image of Aging Scale</i>						
	Levy et al. (2004) USA	Cross-sectional	717 (64.16% female) 10 (60% female)	65 years or older	18 bipolar adjectives	– Content validity – Internal consistency – Reliability – Convergent validity	NA
18	<i>Prescriptive Intergenerational-Tension Ageism Scale</i>						
	North and Fiske (2013) USA	Cross-sectional	427 (61.82% female) 93 (74.19% female)	32.9/25.11	20-item	– Structural validity – Internal consistency – Criterion validity – Measurement invariance – Convergent validity	NA
19	<i>Ambivalent Ageism Scale</i>						
	Cary et al. (2017)/ Canada	Cross-sectional	397 (67% female) 194 (55.7% female) 161 (59% female) 23 (65.22% female)	18.91 ± 1.9 26.9 ± 4.55 25.02 ± 6.57 19.3 ± 3.87	13-item	– Content validity – Structural validity – Internal consistency – Reliability – Criterion validity – Convergent validity	NA
20	<i>Adolescents' Ageism Toward Older Adults Scale</i>						
	Marchetti et al. (2022) Italy	Cross-sectional	575 (57.4% female)	14.6 ± 2.37	33-item	– Content validity – Structural validity – Internal consistency – Reliability – Convergent validity	NA

Abbreviations: NA, Not available; PROM, Patient-Reported Outcome Measure.

TABLE 2 | Summary of content validity, structural validity and internal consistency of ageism instruments.

Author(s)	Content validity (Box 2)	COSMIN score/Quality score	Structural validity (Box 3)	COSMIN score/Quality score ^a	Internal consistency (Box 4)	COSMIN score/Quality score
<i>Philadelphia Geriatric Center (PGC) Morale Scale</i>						
Lawton (1975)	NA	NA	43% of the variance explained by the extracted factors	Adequate/–	Cronbach α : 0.81–0.85	Very good/+
<i>Anxiety about Aging Scale (AAS)</i>						
Lasher and Faulkender (1993)	Asking students about their attitude to ageing	Adequate/–	50.6% of the variance explained by the extracted factors	Very good/–	Cronbach α : 0.82	Very good/+
<i>Expectations Regarding Aging-12 item (ERA-12)</i>						
Sarkisian et al. (2005)	Asking older adults and physicians about expectations regarding ageing	Adequate/–	ERA-12 captures 88% of the variation in the ERA-38 overall score	Very good/–	Cronbach α : 0.89	Very good/+
<i>Aging Perception Questionnaire (APQ)</i>						
Barker et al. (2007)	Relevant population involved	Adequate/+	$\chi^2 = 2788$, $DF = 496$, $CFI = 0.91$, $IFI = 0.91$, $GFI = 0.91$, $RMSEA = 0.05$, $p < 0.001$	Very good/–	Cronbach α : 0.64–0.89	Very good/–
<i>Attitudes to Aging Questionnaire (AAQ)</i>						
Laidlaw et al. (2007)	Relevant population involved	Adequate/+	$CFI = 0.700$	Very good/–	Cronbach α : 0.86	Very good/+
<i>Age-Based Rejection Sensitivity Questionnaire (RSQ-Age)</i>						
Kang and Chasteen (2009)	Asking older adults about reactions to age-based rejection	Adequate/–	43% of the variance explained by the extracted factors.	Adequate/–	Cronbach α : 0.91	Very good/+
<i>Relational Aging Anxiety Scale (RAA)</i>						
Gendron et al. (2020)	Asking older adults about ageing anxiety	Adequate/–	59.68% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.88	Very good/+

(Continues)

TABLE 2 | (Continued)

Author(s)	Content validity (Box 2)	COSMIN score/Quality score	Structural validity (Box 3)	COSMIN score/Quality score ^a	Internal consistency (Box 4)	COSMIN score/Quality score
<i>Everyday Ageism Scale</i>						
Allen et al. (2022)	Asking older adults and professionals about ageism and its relationships with health	Adequate/–	62.66% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.768	Very good/+
<i>Tuckman–Lorge Attitudes Toward Old People</i>						
Axelrod and Eisdorfer (1961)	Asking students about attitudes toward older people	Adequate/–	NA	NA	NA	NA
<i>Kogan Attitudes Toward Old People Scale</i>						
Kogan (1961)	Asking students about attitudes toward older people	Adequate/–	35% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.64–0.69	Very good/–
<i>Aging Semantic Differential</i>						
Rosencranz and McNevin (1969)	NA	NA	NA	NA	NA	NA
<i>Aging Opinion Survey</i>						
Kafer et al. (1980)	Asking students about attitudes towards older people	Adequate/?	24.9% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.60–0.78	Very good/–
<i>Facts on Aging Quiz</i>						
Palmore (1980)	NA	NA	NA	NA	Cronbach α : 0.47	Very good/–
<i>Fraboni Scale of Ageism</i>						
Fraboni et al. (1990)	NA	NA	37.5% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.88	Very good/+

(Continues)

TABLE 2 | (Continued)

Author(s)	Content validity (Box 2)	COSMIN score/Quality score	Structural validity (Box 3)	COSMIN score/Quality score ^a	Internal consistency (Box 4)	COSMIN score/Quality score
<i>Braithwaite Scale of Ageism</i>						
Braithwaite et al. (1993)	Asking students about attitudes towards older people	Adequate/–	NA	NA	Cronbach α : 0.02–0.88	Very good/–
<i>Reactions to Ageing Questionnaire (RAQ)</i>						
Gething (1994)	Asking nurses about attitudes towards older people	Adequate/–	48.7% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.81	Very good/+
<i>Image of Aging Scale</i>						
Levy et al. (2004)	Relevant population involved	Adequate/+	NA	NA	Cronbach α : 0.82–0.84	Very good/+
<i>Prescriptive Intergenerational-Tension Ageism Scale</i>						
North and Fiske (2013)	Asking people about prescriptive beliefs concerning potential intergenerational tensions	Adequate/–	46.51% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.91	Very good/+
<i>Ambivalent Ageism Scale</i>						
Cary et al. (2017)	Asking students about benevolent and hostile attitudes towards older adults	Adequate/–	61.33% of the variance explained by the extracted factors.	Very good/–	Cronbach α : 0.91	Very good/+
<i>Adolescents' Ageism Toward Older Adults Scale</i>						
Marchetti et al. (2022)	Relevant population involved	Adequate/+	Explained variance not mentioned	Very good/?	Cronbach α : 0.80	Very good/+

Abbreviations: ICC, Intra-class correlation coefficient; NA, Not available; r, Pearson's correlation coefficients.
^aQuality score of the measurement property: (+) positive measurement property; (–) negative measurement property; (?) indeterminate.

TABLE 3 | Summary of cross-cultural validity, reliability, convergent validity and known-groups validity of ageism instruments.

Author(s)	Cross-cultural validity (Box 5)	COSMIN score/Quality score	Reliability (Box 6)	COSMIN score/Quality score ^a	Convergent validity (Box 9a)	COSMIN score/Quality score	Known-groups validity (Box 9b)	COSMIN score/Quality score
<i>Philadelphia Geriatric Center (PGC) Morale Scale</i>								
Lawton (1975)	NA	NA	NA	NA	NA	NA	NA	NA
<i>Anxiety about Aging Scale (AAS)</i>								
Lasher and Faulkender (1993)	NA	NA	NA	NA	– Self-Efficacy = –0.516** – Facts on Aging = –0.193**	Very good/+	Age*, gender*, contact**, quality contact**	Very good/+
<i>Expectations Regarding Aging-12 item (ERA-12)</i>								
Sarkisian et al. (2005)	NA	NA	NA	NA	– 12-item Short-Form Health Survey = 0.21–0.28** – Depression symptom = –0.26**	Very good/+	Age**, depressive symptoms**	Very good/+
<i>Aging Perception Questionnaire (APQ)</i>								
Barker et al. (2007)	NA	NA	ICC: 0.76 (Test–retest)	Inadequate/+	– Participant experience of health-related changes – Functional disability – Depression	Very good/+	Age**, gender**, education**, comorbidity**, income	Very good/+
<i>Attitudes to Aging Questionnaire (AAQ)</i>								
Laidlaw et al. (2007)	– Gender – Age (less than 80 years, or equal or greater than 80 years) – Centre (European versus non-European centers)	Very good/+	ICC: 0.16 (Test–retest) ICC: 0.10 (Inter-rater)	Very good/–	NA	NA	NA	NA

(Continues)

TABLE 3 | (Continued)

Author(s)	Cross-cultural validity (Box 5)	COSMIN score/Quality score	Reliability (Box 6)	COSMIN score/Quality score ^a	Convergent validity (Box 9a)	COSMIN score/Quality score	Known-groups validity (Box 9b)	COSMIN score/Quality score
<i>Age-Based Rejection Sensitivity Questionnaire (RSQ-Age)</i>								
Kang and Chasteen (2009)	NA	NA	$r = 0.74^*$ (1-year test-retest)	Adequate/–	– RSQ-Personal (assesses anxious expectations of rejection) = -0.36^{**} – Stigma consciousness scale = -0.27^{**} – Modern sexism scale = 0.06 – Self-consciousness scale = -0.31^{**} – Rosenberg self-esteem scale = -0.49^{**} – Awareness of ageism scale = -0.34^{**}	Very good/+	Age, gender, education level, ethnicity, marital status and physical health*	Adequate/+
<i>Relational Aging Anxiety Scale (RAA)</i>								
Gendron et al. (2020)	NA	NA	NA	NA	NA	NA	Age**, gender*, ethnicity The three variables together accounted for 25% of the variance	Very good/+
<i>Everyday Ageism Scale</i>								
Allen et al. (2022)	NA	NA	NA	NA	NA	NA	Chronological age**, relative appearance**	Very good/+
<i>Tuckman–Lorge's attitudes toward old people</i>								
Axelrod and Eisdorfer (1961)	NA	NA	$r = 0.36-0.62$ (test-retest)	Adequate/–	NA	NA	Age	Adequate/+

(Continues)

TABLE 3 | (Continued)

Author(s)	Cross-cultural validity (Box 5)	COSMIN score/Quality score	Reliability (Box 6)	COSMIN score/Quality score ^a	Convergent validity (Box 9a)	COSMIN score/Quality score	Known-groups validity (Box 9b)	COSMIN score/Quality score
<i>Kogan's Attitudes Toward Old People Scale</i>								
Kogan (1961)	Psychology students from different universities	Adequate/?	NA	NA	- Authoritarianism scales - Anomie scales - Antiminority and disability scales	Adequate/-	NA	NA
<i>Aging Semantic Differential</i>								
Rosencranz and McNevin (1969)	NA	NA	NA	NA	NA	NA	Comparisons of those with close grandparent contact and those with little or no grandparent contact/with at least one meaningful association with aged persons and those with no meaningful associations/with hospital contact with the aged and those with no hospital contact	Adequate/+
<i>Aging Opinion Survey</i>								
Kafer et al. (1980)	NA	NA	NA	NA	NA	NA	Age**, sex, contact ratings	Adequate/+
<i>Facts on Aging Quiz</i>								
Palmore (1980)	NA	NA	NA	NA	- Correlation with Kogan's old people scale = 0.31 - Correlation with the Ageing semantic differential scale = 0.27	Adequate/+	NA	NA

(Continues)

TABLE 3 | (Continued)

Author(s)	Cross-cultural validity (Box 5)	COSMIN score/Quality score	Reliability (Box 6)	COSMIN score/Quality score ^a	Convergent validity (Box 9a)	COSMIN score/Quality score	Known-groups validity (Box 9b)	COSMIN score/Quality score
<i>Fraboni Scale of Ageism</i>								
Fraboni et al. (1990)	NA	NA	NA	NA	- Acceptance, of other's scale = 0.40** - Facts on ageing quiz = -0.28** - Social desirability scale = -0.03	Adequate/+	Age, sex*, socioeconomic status	Adequate/+
<i>Braithwaite Scale of ageism</i>								
Braithwaite et al. (1993)	NA	NA	NA	NA	NA	NA	Age*, interests (research/teaching), contact with older people (low/high)**	Adequate/+
<i>Reactions to Ageing Questionnaire (RAQ)</i>								
Gething (1994)	NA	NA	NA	NA	- Facts on aging quiz = 0.13** - Ageing semantic differential - = -0.26**	Adequate/+	Age*, gender, education, frequency of contact with older people at work, frequency of contact outside work	Very good/+
<i>Image of Aging Scale</i>								
Levy et al. (2004)	NA	NA	$r = 0.79-0.92$ (test-retest)	Adequate/+	NA	NA	Negative age-stereotype component of the scale was associated with greater lifetime-exposure to television*	Adequate/+

(Continues)

TABLE 3 | (Continued)

Author(s)	Cross-cultural validity (Box 5)	COSMIN score/Quality score	Reliability (Box 6)	COSMIN score/Quality score ^a	Convergent validity (Box 9a)	COSMIN score/Quality score	Known-groups validity (Box 9b)	COSMIN score/Quality score
<i>Prescriptive Intergenerational-Tension Ageism Scale</i>								
North and Fiske (2013)	East Asians scoring significantly higher than Whites and Blacks on all three subscales. South Asians outpacing Whites and Blacks on the succession and consumption subscales, and Latinos similarly scoring higher than Blacks and Whites on the succession subscale and lower than East Asians on the consumption subscale. (all p -values < 0.05)	Adequate/+	NA	NA	– Fraboni scale of ageism = 0.86 – Ambivalent sexism inventory = 0.93 – Symbolic racism scale = 0.88 – Social dominance orientation scale = 0.93 – Right-Wing authoritarianism scale = 0.94	Adequate/+	Age**, gender**	Adequate/+
<i>Ambivalent Ageism Scale</i>								
Cary et al. (2017)	NA	NA	$r = 0.76 - 0.80$ (test-retest)	Adequate/+	– Fraboni scale of ageism** – Warmth and Competence stereotypes scale**	Very good/+	NA	NA
<i>Adolescents' Ageism Toward Older Adults Scale (AGES)</i>								
Marchetti et al. (2022)	NA	NA	ICC: 0.71 (Test-retest)	Very good/+	– Correlation with Fraboni Scale of Ageism = 0.63	Adequate/+	Age**, gender, frequency of contact, quality of contact**	Very good/+

Abbreviations: ICC, Intra-class correlation coefficient; NA, Not available; r , Pearson's correlation coefficients.^aQuality score of the measurement property: (+) positive measurement property; (–) negative measurement property; (?) indeterminate.** $p < 0.001$; * $p < 0.05$.

TABLE 4 | The Rasch reliability of self-stigma associated with ageism instruments.

Population and study	Rasch reliability				COSMIN score/ Quality score ^a
	Data fit	Differential item functioning (DIF) analysis	ICC for persons/ coefficient of homogeneity of persons	ICC for items/SCOTT	
<i>Attitudes to Aging Questionnaire (AAQ)</i>					
Laidlaw et al. (2007) 15 centers	Three-factor for all 38 items: CFI = 0.781, $\chi^2 = 11,731.0$, df = 659, RMSEA = 0.055 Three-factor for all 24 items: CFI = 0.842, $\chi^2 = 4559.9$, df = 248, RMSEA = 0.056	DIF analyses were carried out with dichotomised variables for Gender, for Age (less than 80 years, or equal or greater than 80 years), and for Centre (European and for Centre (European vs non-European centres)	0.08–0.10/1.00	0.16/0.16	Very good/+

Abbreviations: CFI, Comparative-fit index; ICC, intraclass correlation; RMSEA, Root-mean-square error of approximation; SCOTT, Scott's homogeneity ratio.

^aQuality score of the measurement property: (+) positive measurement property; (–) negative measurement property; (?) indeterminate.

centres necessitated small and unreliable sample sizes using item response theory prerequisites. Consequently, caution is advised when interpreting centre differential item functioning analyses. Methodological robustness was demonstrated in this study (see Table 4). However, best-evidence synthesis showed limited positive evidence to support the reliability (interrater, intra-rater, and test–retest) of instruments assessing ageism (Laidlaw et al. 2007).

3.7 | Validity

3.7.1 | Convergent Validity

Five studies had very good methodological quality (Barker et al. 2007; Cary et al. 2017; Kang and Chasteen 2009; Lasher and Faulkender 1993; Sarkisian et al. 2005), and six studies had adequate methodological quality (Fraboni et al. 1990; Gething 1994; Kogan 1961; Marchetti et al. 2022; North and Fiske 2013; Palmore 1980). Correlations were reported with measures of self-efficacy, self-consciousness, self-esteem, anxiety, depression, quality of life, functional disability, sexism, racism, and authoritarianism. For example, among the studies reporting associations with the Facts on Aging (Palmore 1980) and the Fraboni Scale of Ageism (Fraboni et al. 1990), correlation coefficients ranged from 0.13 to 0.28 and 0.63 to 0.86, respectively. Therefore, the best-evidence synthesis yielded an unknown level of evidence for convergent validity (see Table 5).

3.7.2 | Known-Groups Validity

In distinguishing their target populations, seven studies demonstrated very good methodological quality (Allen et al. 2022; Barker et al. 2007; Gendron et al. 2020; Gething 1994; Lasher and Faulkender 1993; Marchetti et al. 2022; Sarkisian et al. 2005), whereas eight studies displayed adequate methodological quality (Axelrod and Eisdorfer 1961; Braithwaite et al. 1993; Fraboni et al. 1990; Kafer et al. 1980; Kang and Chasteen 2009; Levy et al. 2004; North and Fiske 2013; Rosencranz and McNevin 1969). Ageism-related instruments are generally associated with factors such as age, gender, depressive symptoms and how often individuals interact with older adults. However, they often do not differentiate between specific age groups, such as ageism towards younger men and ageism towards older men.

4 | Discussion

The purpose of the present review was to evaluate the psychometric properties of the instruments used to assess ageism. A total of 20 studies that had developed psychometric instruments to assess ageism (Allen et al. 2022; Axelrod and Eisdorfer 1961; Barker et al. 2007; Braithwaite et al. 1993; Cary et al. 2017; Fraboni et al. 1990; Gendron et al. 2020; Gething 1994; Kafer et al. 1980; Kang and Chasteen 2009; Kogan 1961; Laidlaw et al. 2007; Lasher and Faulkender 1993; Lawton 1975; Levy et al. 2004; Marchetti et al. 2022; North and Fiske 2013; Palmore 1980; Rosencranz and McNevin 1969; Sarkisian et al. 2005) were evaluated to determine the quality of psychometric properties, including content validity, structural validity,

TABLE 5 | Levels of evidence of various measurement properties of the ageism instruments.

Study	Content validity	Structural validity	Internal consistency	Reliability	Cross-cultural validity	Construct validity	
						Known-group	Convergent
Philadelphia Geriatric Center (PGC) Morale Scale	NA	Limited (−)	Strong (+)	NA	NA	NA	NA
Anxiety about Aging Scale (AAS)	Limited (−)	Strong (−)	Strong (+)	NA	NA	Strong (+)	Strong (+)
Expectations Regarding Aging-12 item (ERA-12)	Limited (−)	Strong (−)	Strong (+)	NA	NA	Strong (+)	Strong (+)
Aging Perception Questionnaire (APQ)	Limited (+)	Strong (−)	Strong (+)	NA	Limited (−)	Strong (+)	Strong (+)
Attitudes to Aging Questionnaire (AAQ)	Limited (+)	Strong (−)	Strong (+)	Strong (+)	Strong (−)	NA	NA
Age-Based Rejection Sensitivity Questionnaire (RSQ-Age)	Limited (−)	Limited (+)	Strong (+)	NA	Limited (−)	Strong (+)	Limited (+)
Relational Aging Anxiety Scale (RAA)	Limited (−)	Strong (−)	Strong (+)	NA	NA	NA	Strong (+)
Everyday Ageism Scale	Limited (−)	Strong (−)	Strong (+)	NA	NA	NA	Strong (+)
Tuckman–Lorge's Attitudes Toward Old People	Limited (−)	?	NA	NA	Limited (−)	NA	Limited (+)
Kogan's Attitudes Toward Old People Scale	Limited (−)	Strong (−)	Strong (−)	?	?	Limited (−)	NA
Aging Sematic Differential	NA	NA	NA	NA	NA	NA	Limited (+)
Aging Opinion Survey	?	Strong (−)	Strong (−)	NA	NA	NA	Limited (+)
Facts on Aging Quiz	NA	NA	NA	NA	NA	Limited (+)	NA
Fraboni Scale of Ageism	NA	Strong (−)	Strong (+)	NA	NA	Limited (+)	Limited (+)
Braithwaite Scale of Ageism	Limited (−)	NA	Strong (−)	NA	NA	NA	Limited (+)
Reactions to Ageing Questionnaire (RAQ)	Limited (−)	Strong (−)	Strong (+)	NA	NA	Limited (+)	Strong (+)
Image of Aging Scale	Limited (−)	Strong (−)	Strong (+)	NA	Limited (+)	NA	Limited (+)

(Continues)

TABLE 5 | (Continued)

Study	Content validity	Structural validity	Internal consistency	Reliability	Cross-cultural validity	Construct validity	
						Known-group	Convergent
Prescriptive Intergenerational-Tension Ageism Scale	Limited (−)	Strong (−)	Strong (+)	Limited (+)	NA	Limited (+)	Limited (+)
Ambivalent Ageism Scale	Limited (−)	Strong (−)	Strong (+)	NA	Limited (+)	Strong (+)	NA
Adolescents' Ageism Toward Older Adults Scale (AGES)	Limited (+)	?	Strong (+)	NA	Strong (+)	Limited (−)	Strong (+)

Note: (+) positive result rating; (−) negative result rating; (±) conflicting; (?) unknown. Positive result ratings (+) signify that the measurement property satisfies the quality criteria outlined in Appendix S2. For instance, internal consistency is met when Cronbach's α falls between 0.70 and 0.95, or structural validity is achieved when factors explain more than 50% of the variance. Conversely, negative result ratings (−) indicate non-fulfilment of these criteria, such as Cronbach's α below 0.70 or variance explained below 50%. Evident measurement properties are represented by (+) or conflicting (+), reflecting tested properties with at least partial evidence. Non-evident properties correspond to (−) or (?) where evidence is insufficient or indeterminate.

cross-cultural validity, internal consistency, reliability, convergent validity, and known-groups validity. Internal consistency was examined in 18 studies (Allen et al. 2022; Barker et al. 2007; Braithwaite et al. 1993; Cary et al. 2017; Fraboni et al. 1990; Gendron et al. 2020; Gething 1994; Kafer et al. 1980; Kang and Chasteen 2009; Kogan 1961; Laidlaw et al. 2007; Lasher and Faulkender 1993; Lawton 1975; Levy et al. 2004; Marchetti et al. 2022; North and Fiske 2013; Palmore 1980; Sarkisian et al. 2005), while cross-cultural validity was assessed in only three (Kogan 1961; Laidlaw et al. 2007; North and Fiske 2013). Among these, six studies (i.e. Marchetti et al. 2022; Barker et al. 2007; Laidlaw et al. 2007; Lasher and Faulkender 1993; Sarkisian et al. 2005) were considered to have very good methodological quality because they had four or more psychometric properties with very good quality (i.e. over half of the evaluated psychometric properties because the present study evaluated eight properties). However, none provided sufficient evidence to deem their content validity or reliability as 'good'. Importantly, none of the instruments were classified as inadequate based on their worst score counts, highlighting that while improvements are needed, existing instruments exhibit baseline psychometric soundness.

The present review went much further than Ayalon et al. (2019) because it incorporated nine additional instruments, not in the 2019 review, including the Philadelphia Geriatric Center (PGC) Morale Scale (Lawton 1975) and the Ambivalent Ageism Scale (Cary et al. 2017), among others. While Ayalon et al.'s review focused primarily on established instruments, the present study expanded the scope by identifying gaps, such as insufficient cross-cultural validation and the limited focus on discrimination behaviours. Additionally, the present review identified divergences in methodological quality by adhering more rigorously to the COSMIN standards. However, both Ayalon et al. and the present authors emphasise the need for more comprehensive and psychometrically robust instruments.

Content validity emerged as a critical factor in determining the adequacy of the psychometric instruments, given its essential

role in ensuring relevance and comprehensiveness. Many studies lacked meaningful engagement with target populations during item generation, which represents a significant limitation in establishing validity. As emphasised in the COSMIN guidelines, content validity serves as the foundation for all subsequent psychometric evaluations, underscoring the necessity of its prioritisation (Terwee et al. 2018). Despite its importance, none of the studies included in the present review received a high methodological quality rating for content validity based on COSMIN criteria (Table 2).

The majority of studies focused on older adults and students, employing cross-sectional designs for data collection. Iversen et al. (2009) highlighted the need to recognise ageism in its diverse manifestations—positive and negative, explicit and implicit—operating across individual, interpersonal, and institutional levels. Existing psychometric scales in ageism research primarily assess ageing-related prejudices and stereotypes or attitudes towards older adults. However, very few instruments assess active endorsement or engagement in discriminatory behaviours against older adults. Moreover, there is a notable absence of instruments that evaluate the societal impacts of ageist stereotypes, prejudices, and discriminatory actions on older populations. Additionally, two studies (Barker et al. 2007; Sarkisian et al. 2005) reported that self-directed ageism was positively associated with depression ($r=0.667$, and $r=0.33$, respectively). These findings underscore the need for more comprehensive instruments capable of capturing the multifaceted nature of ageism and its broader implications.

Based on the findings of the present review, the currently available measures of ageism require further development to address significant gaps in their scope and applicability. While some psychometric scales assess older adults' reactions to being stereotyped (Barker et al. 2007; Levy et al. 2004; Palmore 1980; Rosencranz and McNevin 1969; Sarkisian et al. 2005; Tuckman and Lorge 1952), these instruments often lack comprehensiveness. Many existing instruments focus narrowly on specific aspects of ageism (e.g. explicit attitudes or stereotypes) without

capturing its broader dimensions, including implicit biases, systemic discrimination, and the intersectionality of ageism with other forms of prejudice. Moreover, few scales evaluate the full range of behaviours and consequences associated with ageism, such as discriminatory practices, their psychological and social impacts on older adults, and the influence of ageist beliefs on policy and institutional structures. These limitations hinder the ability to fully understand and address ageism as a multifaceted phenomenon.

To develop more robust instruments, a comprehensive and reliable set of indicators is needed. These indicators should assess explicit and implicit ageist attitudes, discriminatory behaviours, the impact of ageism on individuals' mental and physical health, and the structural and institutional manifestations of ageism. Such instruments should also account for cultural and demographic variations, enabling their application across diverse populations and settings. Developing these advanced measures would provide a more nuanced understanding of ageism and support the design of targeted interventions to mitigate its pervasive effects.

Ayalon et al.'s review provided a foundational assessment of the psychometric properties of ageism instruments published before 2017, focusing on a narrower range of instruments and psychometric dimensions. Building on this study, the present review broadened the scope by including literature from 2017 to 2024, expanding the range of instruments examined, and offering a detailed critique across additional psychometric properties such as structural validity, cross-cultural validity, and reliability. This extended scope not only underscores critical gaps in existing measures, such as limited content validity and inadequate cross-cultural validation, but also highlights areas requiring refinement and innovation in instrument development.

Furthermore, the present study emphasises specific dimensions of ageism—self-directed, interpersonal and institutional—to provide a more comprehensive understanding of the construct. By examining how ageism manifests at individual, relational, and systemic levels, the review advances beyond a generalised perspective, enabling a nuanced critique of existing instruments. This multifaceted approach ensures that the instruments assessed can effectively capture these distinct dimensions, while also identifying gaps that need to be addressed in future research and instrument development.

The review's primary strength lies in its methodical search strategy, which adhered to the COSMIN checklist and quality criteria to ensure a rigorous evaluation of the included studies. However, specific methodological considerations warrant discussion. First, the exclusion of grey literature, including studies identified through search engines such as *Google* and *Google Scholar*, may have narrowed the range of studies reviewed. However, this decision was made to prioritise methodological rigour and reliability, and it is unlikely to have significantly affected the overall findings. Second, the diversity of instruments reviewed limits direct comparisons between instruments, and as such, no single instrument was identified as the most robust or preferred. Instead, the review offers critical insights for advancing the field, emphasising the need for standardisation and

further validation of ageism instruments to enhance their applicability across diverse populations and contexts.

5 | Conclusion

The present review provides a comprehensive synthesis of multi-component ageism instruments with published psychometric data, critically evaluating their quality and scope. Despite identifying 20 instruments, significant gaps remain regarding the comprehensiveness of their psychometric properties, including limited methodological rigour and variability in the constructs assessed. The findings underscore the lack of consideration regarding the developmental trajectory for assessing ageism, particularly in clinical and community-based settings, which complicates efforts in evaluating the construct consistently across diverse populations. Understanding the developmental trajectory of ageism across the lifespan is critical for informing interventions aimed at reducing its prevalence and mitigating its impact on the ageing population. Future research should prioritise refining existing instruments and developing psychometrically robust instruments tailored to diverse contexts and settings.

Author Contributions

All authors have made an active contribution to the conception and design, acquisition of data, analysis and interpretation of data.

Acknowledgements

The present study was supported by funding from the National Health Research Institutes, Taiwan (NHRI-11A1-CG-CO-04-2225-1) and Academia Sinica, Taiwan (AS-GCS-113-H01).

Disclosure

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

- Allen, J. O., E. Solway, M. Kirch, D. Singer, J. T. Kullgren, and P. N. Malani. 2022. "The Everyday Ageism Scale: Development and Evaluation." *Journal of Aging and Health* 34, no. 2: 147–157. <https://doi.org/10.1177/08982643211036131>.
- Axelrod, S., and C. Eisdorfer. 1961. "Attitudes Toward Old People: An Empirical Analysis of the Stimulus-Group Validity of the

- Tuckman-Lorge Questionnaire." *Journal of Gerontology* 16: 75–80. <https://doi.org/10.1093/geronj/16.1.75>.
- Ayalon, L., P. Dolberg, S. Mikulionienė, et al. 2019. "A Systematic Review of Existing Ageism Scales." *Ageing Research Reviews* 54: 100919. <https://doi.org/10.1016/j.arr.2019.100919>.
- Barker, M., A. O'Hanlon, H. M. McGee, A. Hickey, and R. M. Conroy. 2007. "Cross-Sectional Validation of the Aging Perceptions Questionnaire: A Multidimensional Instrument for Assessing Self-Perceptions of Aging." *BMC Geriatrics* 7: 9. <https://doi.org/10.1186/1471-2318-7-9>.
- Ben-Harush, A., S. Shiovitz-Ezra, I. Doron, et al. 2017. "Ageism Among Physicians, Nurses, and Social Workers: Findings From a Qualitative Study." *European Journal of Ageing* 14: 39–48.
- Braithwaite, V., R. Lynd-Stevenson, and D. Pigram. 1993. "An Empirical Study of Ageism: From Polemics to Scientific Utility." *Australian Psychologist* 28, no. 1: 9–15. <https://doi.org/10.1080/00050069308258857>.
- Butler, R. N. 1969. "Age-Is: Another Form of Bigotry." *Gerontologist* 9, no. 4_Part_1: 243–246. https://doi.org/10.1093/geront/9.4_Part_1.243.
- Cary, L. A., A. L. Chasteen, and J. Remedios. 2017. "The Ambivalent Ageism Scale: Developing and Validating a Scale to Measure Benevolent and Hostile Ageism." *Gerontologist* 57, no. 2: e27–e36. <https://doi.org/10.1093/geront/gnw118>.
- Chen, Y.-J., S. Kukreti, H.-L. Yang, et al. 2023. "Psychometric Properties of Instruments Assessing Intrinsic Capacity: A Systematic Review." *Asian Journal of Social Health and Behavior* 6, no. 4: 141–155.
- Cherry, K. E., and E. Palmore. 2008. "Relating to Older People Evaluation (ROPE): A Measure of Self-Reported Ageism." *Educational Gerontology* 34, no. 10: 849–861. <https://doi.org/10.1080/03601270802042099>.
- de São José, J. M. S., and C. A. F. Amado. 2017. "On Studying Ageism in Long-Term Care: A Systematic Review of the Literature." *International Psychogeriatrics* 29, no. 3: 373–387.
- Fraboni, M., R. Saltstone, and S. Hughes. 1990. "The Fraboni Scale of Ageism (FSA): An Attempt at a More Precise Measure of Ageism." *Canadian Journal on Aging/La Revue Canadienne du Vieillessement* 9, no. 1: 56–66. <https://doi.org/10.1017/S0714980800016093>.
- Gebremariam, K. M., and R. Sadana. 2019. "On the Ethics of Healthy Ageing: Setting Impermissible Trade-Offs Relating to the Health and Well-Being of Older Adults on the Path to Universal Health Coverage." *International Journal for Equity in Health* 18: 1–16.
- Gendron, T., J. K. Inker, R. Andricosky, and F. Zanjani. 2020. "Development of the Relational Ageism Scale: Confirmatory Test on Survey Data." *International Journal of Aging and Human Development* 90, no. 3: 281–296. <https://doi.org/10.1177/0091415019836956>.
- Gething, L. 1994. "Health Professional Attitudes Towards Ageing and Older People: Preliminary Report of the Reactions to Ageing Questionnaire." *Australasian Journal on Ageing* 13, no. 2: 77–81. <https://doi.org/10.1111/j.1741-6612.1994.tb00646.x>.
- Hu, R. X., M. Luo, A. Zhang, and L. W. Li. 2021. "Associations of Ageism and Health: A Systematic Review of Quantitative Observational Studies." *Research on Aging* 43, no. 7–8: 311–322. <https://doi.org/10.1177/0164027520980130>.
- Iversen, T. N., L. Larsen, and P. E. Solem. 2009. "A Conceptual Analysis of Ageism." *Nordic Psychology* 61, no. 3: 4–22. <https://doi.org/10.1027/1901-2276.61.3.4>.
- Kafer, R. A., W. Rakowski, M. Lachman, and T. Hickey. 1980. "Aging Opinion Survey: A Report on Instrument Development." *International Journal of Aging and Human Development* 11, no. 4: 319–333. <https://doi.org/10.2190/JQF5-XDCV-H1AH-3E1Y>.
- Kang, S. K., and A. L. Chasteen. 2009. "The Development and Validation of the Age-Based Rejection Sensitivity Questionnaire." *Gerontologist* 49, no. 3: 303–316. <https://doi.org/10.1093/geront/gnp035>.
- Kogan, N. 1961. "Attitudes Toward Old People: The Development of a Scale and an Examination of Correlates." *Journal of Abnormal and Social Psychology* 62, no. 1: 44–54. <https://doi.org/10.1037/h0048053>.
- Laidlaw, K., M. J. Power, and S. Schmidt. 2007. "The Attitudes to Ageing Questionnaire (AAQ): Development and Psychometric Properties." *International Journal of Geriatric Psychiatry* 22, no. 4: 367–379. <https://doi.org/10.1002/gps.1683>.
- Lasher, K. P., and P. J. Faulkender. 1993. "Measurement of Aging Anxiety: Development of the Anxiety About Aging Scale." *International Journal of Aging and Human Development* 37, no. 4: 247–259. <https://doi.org/10.2190/1u69-9au2-v6lh-9y1l>.
- Lawton, M. P. 1975. "The Philadelphia Geriatric Center Morale Scale: A Revision." *Journal of Gerontology* 30, no. 1: 85–89. <https://doi.org/10.1093/geronj/30.1.85>.
- Lee, C. Y., H. C. Hsu, Y. L. Chou, and Y. J. Chen. 2012. "Willingness and Attitudes Toward Elderly Care Among Nurses at a Medical Center in Central Taiwan." *Journal of Nursing and Healthcare Research* 8, no. 4: 297–305. <https://doi.org/10.6225/JNHR.08.4.297>.
- Levy, B. R., S. V. Kasl, and T. M. Gill. 2004. "Image of Aging Scale." *Perceptual and Motor Skills* 99, no. 1: 208–210. <https://doi.org/10.2466/PMS.99.4.208-210>.
- Marchetti, A., M. Lommi, C. Barbaranelli, M. Piredda, M. G. De Marinis, and M. Matarese. 2022. "Development and Initial Validation of the Adolescents' Ageism Toward Older Adults Scale." *Gerontologist* 62, no. 3: e150–e161. <https://doi.org/10.1093/geront/gnab023>.
- Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *Annals of Internal Medicine* 151, no. 4: 264–269.
- North, M. S., and S. T. Fiske. 2013. "A Prescriptive Intergenerational-Tension Ageism Scale: Succession, Identity, and Consumption (SIC)." *Psychological Assessment* 25, no. 3: 706–713. <https://doi.org/10.1037/a0032367>.
- Palmore, E. 1980. "The Facts on Aging Quiz: A Review of Findings." *Gerontologist* 20, no. 6: 669–672. <https://doi.org/10.1093/geront/20.6.669>.
- Prinsen, C. A., S. Vohra, M. R. Rose, et al. 2016. "How to Select Outcome Measurement Instruments for Outcomes Included in a "Core Outcome Set" – A Practical Guideline." *Trials* 17, no. 1: 449. <https://doi.org/10.1186/s13063-016-1555-2>.
- Rosencranz, H. A., and T. E. McNevin. 1969. "A Factor Analysis of Attitudes Toward the Aged." *Gerontologist* 9, no. 1: 55–59. <https://doi.org/10.1093/geront/9.1.55>.
- Sarkisian, C. A., W. N. Steers, R. D. Hays, and C. M. Mangione. 2005. "Development of the 12-Item Expectations Regarding Aging Survey." *Gerontologist* 45, no. 2: 240–248. <https://doi.org/10.1093/geront/45.2.240>.
- Smith, M. L., C. D. Bergeron, C. Cowart, et al. 2017. "Factors Associated With Ageist Attitudes Among College Students." *Geriatrics & Gerontology International* 17, no. 10: 1698–1706. <https://doi.org/10.1111/ggi.12894>.
- Terwee, C. B., S. D. Bot, M. R. de Boer, et al. 2007. "Quality Criteria Were Proposed for Measurement Properties of Health Status Questionnaires." *Journal of Clinical Epidemiology* 60, no. 1: 34–42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>.
- Terwee, C. B., E. P. Jansma, I. I. Riphagen, and H. C. de Vet. 2009. "Development of a Methodological PubMed Search Filter for Finding Studies on Measurement Properties of Measurement Instruments." *Quality of Life Research* 18: 1115–1123.
- Terwee, C. B., L. B. Mokkink, D. L. Knol, R. W. Ostelo, L. M. Bouter, and H. C. de Vet. 2012. "Rating the Methodological Quality in Systematic Reviews of Studies on Measurement Properties: A Scoring System for

the COSMIN Checklist.” *Quality of Life Research* 21, no. 4: 651–657. <https://doi.org/10.1007/s11136-011-9960-1>.

Terwee, C. B., C. A. C. Prinsen, A. Chiarotto, et al. 2018. “COSMIN Methodology for Evaluating the Content Validity of Patient-Reported Outcome Measures: A Delphi Study.” *Quality of Life Research* 27, no. 5: 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>.

Tuckman, J., and I. Lorge. 1952. “Attitudes Toward Older Workers.” *Journal of Applied Psychology* 36, no. 3: 149–153. <https://doi.org/10.1037/h0061650>.

Tuckman, J., and I. Lorge. 1958. “Attitude Toward Aging of Individuals With Experiences With the Aged.” *Journal of Genetic Psychology* 92, no. 2: 199–204. <https://doi.org/10.1080/00221325.1958.10532393>.

United Nations, Department of Economic and Social Affairs, Population Division. 2017. “World Population Ageing 2017 (ST/ESA/SER.A/408).”

World Health Organization. 2021. “Global Report on Ageism.” <https://www.who.int/publications-detail-redirect/9789240016866>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.