

Ethics and Language Assessment Standards in Sociocultural-Historical Contexts: A Systematic Review

Thaweesak Chanpradit,^{1*} Erikson Saragih²

Article History:

Received: 15/04/2026

Revised : 05/06/2026

Accepted: 15/06/2026


Available Online: 30/06/2026


Keywords:

accountability, ethics and language assessment standards, fairness

ABSTRACT

Ethics and language assessment standards ensure assessors' fairness, accountability, and transparency. However, these can be questionable in different sociocultural historical contexts. This review seeks an understanding of how ethics and standards intersect with cultural-historical factors in language assessment, including how assessor bias could be minimized in different cultural-historical contexts. Herein, the review examines research articles on ethical considerations in assessment standards, cultural responsiveness, historical developments, impacts on assessment outcomes, and challenges and solutions. Using the PRISMA procedure with citation chaining, 40 peer-reviewed articles were identified and included in the review. The findings revealed the rise of justice-oriented and culturally sustaining ethical frameworks that enhance fairness in language assessment but remain difficult to fully implement in actual ELT practices. Cultural responsiveness tends to be restricted by standardization, institutional constraints, and colonial legacies that continue to shape assessment ethics. However, when applied in an effective manner, ethical standards could improve fairness and stakeholder trust, yet systemic inequities, limited resources, and emerging issues from AI integration would pose ongoing challenges. Therefore, policymakers, educators, and test developers should stress the need for collaborative, context-sensitive, and reflective approaches that further promote culturally grounded, socially just, and technologically informed ethical practices in language assessment.

¹ Kasetsart University, Bangkok, Thailand. Email: taweesak.j@ku.th  <https://orcid.org/0009-0003-0992-3119> *Corresponding author

² Universitas Sumatera Utara, Medan, Indonesia. Email: eriksonsaragih@usu.ac.id
 <https://orcid.org/0000-0003-0679-6639>

INTRODUCTION

Ethics and standards in language assessment have emerged as a critical issue among educators and researchers, due to the growing recognition of the sociocultural implications of language testing practices. To date, language assessment and testing have evolved from psychometric properties to broader ethical and social justice concerns, reflecting shifts in theoretical perspectives and assessment methodologies (Bachman, 2024; Cervatiuc, 2024).

Ethics and standards in language assessment are at the core of “appraising or estimating the level or magnitude of some attribute of a person” (Mousavi, 2009, p. 35 as cited in Brown & Abeywickrama, 2018, p. 3), especially high-stakes contexts in relation to education, employment, and citizenship, where inequities and biases can have consequences more or less (Schissel, 2023; Shohamy, 1997). For example, standardized tests like IELTS and TOEFL are designed and administered by organizations in Western countries, which often neglect the linguistic and sociocultural diversity of those testers (Baidoo-Anu et al., 2023; Ojochegbe, 2024). This calls for culturally responsive assessment models that further highlight the practical urgency of addressing ethical standards in language assessment (Ghanbari, 2024; Walker et al., 2023). However, significant challenges remain in aligning ethical considerations with language assessment standards, particularly regarding cultural influences and historical developments.

Prior studies (e.g., Baidoo-Anu et al., 2023; Liu et al., 2024; Randall et al., 2023; Walker et al., 2023) have revealed gaps in how ethical frameworks account for the sociocultural contexts of test-takers and the impact of systemic inequities embedded in assessment design and implementation. Additionally, controversies persist between traditional psychometric perspectives emphasizing reliability and validity and critical perspectives advocating for justice-oriented, antiracist validation and culturally sustaining assessment practices (Englert & Shultz, 2025; Plakans & Lee, 2025; Randall et al., 2023). However, educational inequities and the unfairness and invalidity of assessment outcomes still exist in language assessment and testing as long as human beings are concerned (Schissel, 2023). Language assessment involves human decisions, social contexts, and power dynamics. The issue for educators and test makers is not how to achieve full equity; rather, it is how an awareness of inequity and a reduction of biases can be promoted. Hence, it is important to ensure that ethics and standards in language assessment encompass fairness, justice, and responsibility, while cultural responsiveness involves adapting assessments to reflect diverse linguistic and cultural backgrounds (Liu et al., 2024; Wang, 2024).

In light of these gaps, it is essential to better understand how ethics and standards interconnect with cultural-historical factors in language assessment and how bias is reduced in different cultural-historical contexts. Therefore, this review seeks to examine research articles on ethical considerations in language assessment standards. The emphasis is placed on ethical frameworks, cultural responsiveness integration, historical evolution of language assessment standards, impact on language assessment, and implementation challenges and solutions. This review is addressed to the following research questions:

- 1) How are ethics and language assessment standards conceptualized and operationalized in different sociocultural-historical contexts?
- 2) How does cultural responsiveness influence the design, implementation, and outcomes of language assessment practices?
- 3) What challenges, solutions, and themes are reported in implementing ethical and culturally responsive language assessment practices?

Given the research questions, this review is expected to contribute to advancing equitable and culturally responsive assessment practices and inform policymakers, educators, and test developers seeking to uphold ethical standards in different sociocultural contexts. This systematic review seeks to provide useful frameworks that exemplify how ethical considerations can be linked to assessment and evaluation. For example, it could include references that focus on critical approaches to thinking about assessment as a social process rather than purely a measurement activity, and the importance of stakeholders and culturally sensitive designs.

METHOD

Research Design

This study employed a systematic review method. This involved the in-depth synthesis of data in meta-analysis, assessing the reliability of research evidence (Brignardello-Petersen & Guyatt, 2025) in the area of ethics and language assessment standards. The focus is on cultural influences, historical developments, and the impact of ethical standards on assessment implementation through selected research studies. Literature searches were conducted using the following databases: ERIC, Scopus, SAGE Online Journal, Google Scholar, and ResearchGate. The searches were set to collect peer-reviewed articles from the predetermined databases. In the literature search process, metadata were obtained using the following keywords: ethics and standards in language assessment, ethical considerations in language assessment standards, cultural influences in language assessment, historical developments in language assessment, and ethical standards on assessment implementation.

Data Collection

Inclusion and Exclusion Criteria

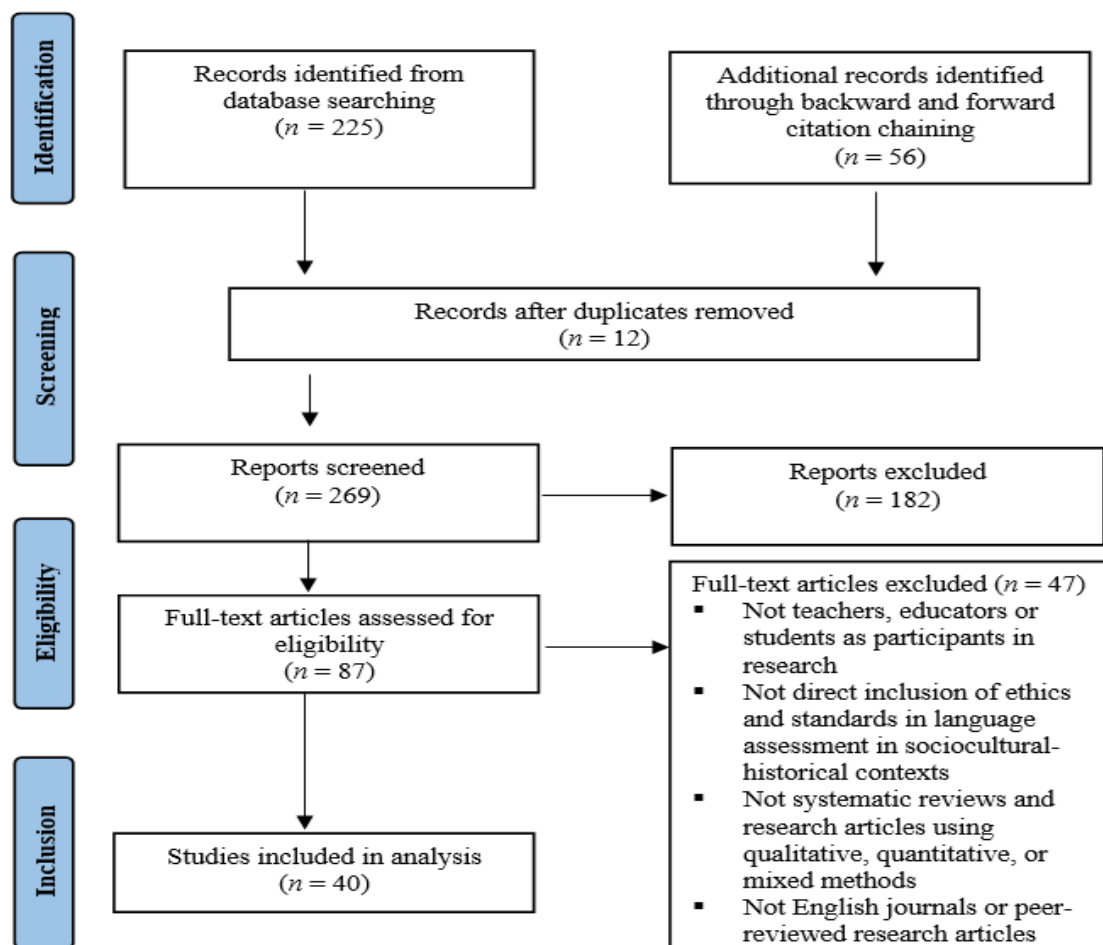
The reviewed studies adhered to the following inclusion and exclusion criteria:

- 1) The studies were focused on teachers, educators, and students who are involved in ethics and standards in language assessment. Studies that involved participants outside of this scope were excluded.
- 2) The studies examined ethics and standards in language assessment. The specific points of the studies included cultural responsiveness integration, historical evolution of standards, impact on assessment outcomes, and implementation challenges and solutions. Studies that were conducted outside of this scope were excluded.
- 3) The studies employed quantitative, qualitative, and mixed-methods approaches to ensure the reliability and validity of the data collection, analysis, and results.
- 4) The studies were published in peer-reviewed academic journals or books to ensure their quality, reliability, and validity. Other forms of literature, such as conference proceedings, book reviews, dissertations, and non-peer-reviewed sources, were excluded.

Screening Procedure

The PRISMA flowchart (Moher et al., 2009), along with backward and forward citation chaining, was adopted in the screening procedure of this review, as illustrated in Figure 1. The PRISMA method is composed of the main stages of identification, screening, eligibility, and inclusion.

Figure 1
PRISMA Flowchart (Moher et al., 2009)



The initial database search identified 225 studies through the use of the keywords mentioned in the search process. To enhance the assurance and quality of the search, additional methods such as backward and forward citation chaining were applied. By reviewing the reference lists of all retrieved articles, 56 studies were added to the initial pool ($n = 225$). Following the removal of 12 duplicate records, 269 studies remained. These studies underwent a two-step screening process. In the first step, titles and abstracts were evaluated for quality using the framework proposed by Jesson et al. (2011), resulting in the selection of 87 studies. The second step involved a full-text eligibility assessment based on the predefined inclusion criteria. From this, 40 studies were included in this review, while 47 studies were excluded, according to the exclusion criteria. To ensure methodological triangulation, citation chaining, both backward and forward, was employed during the initial database search. Backward citation chaining was to ensure that each reference list of the core articles being examined was used to track earlier studies it drew upon. By doing this, previous research works were not overlooked. Meanwhile, forward citation chaining was used to identify newer research works that have cited each core article, tracking how ethics and language assessment standards are being operationalized in sociocultural-historical contexts

that have built on those results. This could uncover emerging debates, replication studies, and recent methodological advances.

Data Analysis

A thematic analysis was guided by the framework of Braun and Clarke (2006). This analysis was undertaken to uncover the central themes within the selected studies. To begin, the studies were categorized into clusters reflecting their research aims and areas of focus. The process commenced with an intensive reading and rereading of the full texts to gain a comprehensive grasp of their content. During this familiarization phase, reflective notes were taken to capture early impressions and potential patterns. Subsequently, significant words, phrases, and sentences that addressed the research questions were systematically coded. These codes were then grouped into meaningful categories, forming the foundation for initial thematic development. Each preliminary theme underwent a careful review to ensure internal consistency, distinctiveness, and accurate representation of the data set. Themes that lacked clarity or sufficient evidence were revised, merged with others, or removed. After refinement, the finalized themes were defined and labeled to reflect their core meaning. Finally, all related codes and illustrative excerpts were compiled into organized tables, which served as the basis for synthesizing and consolidating the final results of the analysis.

RESULTS AND DISCUSSION

Characteristics of the Reviewed Studies

All of the reviewed studies ($n = 40$) adopted one of the following types: qualitative, quantitative, and mixed methods. The studies employed approaches ranging from qualitative case studies and theoretical reviews to quantitative studies and mixed-methods policy analyses. The studies represented a body of literature spanning from the 1990s to 2025, with a surge in publications from 2020 onwards reflecting increased attention to social justice issues in assessment across geographic contexts. These included North America, Europe, Asia, Latin America, and indigenous communities. Collectively, the studies examined ethical frameworks grounded in validation theory, social justice and equity principles, culturally responsive assessment, and professional ethics. The key themes addressed in the studies were fairness and bias reduction, cultural influences on assessment standards, the historical evolution of ethical considerations, and implementation challenges and solutions. Additionally, the reviewed studies focused on different populations—including multilingual learners, indigenous students, and marginalized communities—across various assessment contexts, from large-scale standardized tests to classroom-based evaluations. In a nutshell, Table 1 provides an overview of the reviewed studies focusing on ethical frameworks (EFs), cultural responsiveness integration (CRI), historical evolution (HE) of language assessment standards (LAS), impact on language assessment (IOLA), and implementation challenges and solutions (ICAS).

Table 1
Overview of the Reviewed Studies

No	Author(s) & Date	EFs	CRI	HE of LAS	IOLA	ICAS
1	Randall et al. (2023)	Justice-oriented, antiracist validation framework	Strong focus on marginalized racial groups and systemic bias	Contemporary critique of legacy systems in assessment	Improved fairness through antiracist validation processes	Need for systemic disruption and stakeholder engagement
2	Bachman (2024)	Ethical conduct grounded in validation and professionalization	Recognition of ethical test use as integral to validity	Review of 20 years of ethical and validation developments	Ethical test use is linked to quality control and fairness	Integration of ethics in validation research and practice
3	Ojochegbe (2024)	Ethical critique of standardized tests; advocacy for culturally responsive models	Proposes culturally sensitive frameworks for English proficiency tests	Highlights the limitations of traditional standardized testing	Addresses bias and improves validity for non-Western learners	Framework development for equitable test design
4	Ma (2025)	Theoretical frameworks combining validity, fairness, and epistemic pluralism	Calls for culturally responsive content and modular testing	Historical evolution of CET-4 and its limitations	Identifies inequities and construct-irrelevant variance	Stakeholder participation and equity-based funding models
5	Wang (2024)	Ethical concerns linked to social justice and validity in L2 testing	Contextualized ethics across languages and stakeholders	Under-addressed situated ethics in test validation theory	Social justice and equity are emphasized in assessment validity	Calls for critical approaches and stakeholder voices inclusion
6	Liu et al. (2024)	Ethical judgment differences are influenced by cultural contexts	Cross-cultural comparison of teacher ethical decisions in assessment	Limited historical focus; emphasizes current ethical reasoning	Cultural differences affect fairness perceptions and decisions	Raises awareness of cultural differences in ethical judgment
7	Solano-Flores (2019)	Validity argumentation incorporating cultural procedural assumptions	Cultural responsiveness is embedded in all assessment process components	Systemic and societal efforts are required for cultural responsiveness	Validity arguments integrate cultural fairness evidence	Calls for global efforts to address cultural issues in assessment
8	Baidoo-Anu et al. (2023)	Cultural validity as a core ethical and validity component	Promotes culturally responsive assessment design and decision-making	Challenges include standardization and a lack of resources	Cultural validity enhances fairness and instructional relevance	Professional development and participatory design are needed
9	Steele et al. (2024)	Ethical stance on culturally responsive assessment for Indigenous students	Assessment aligned with culturally sustaining pedagogies	Advocates for proactive validity evaluation from assessment design	Improved validity and fairness for Indigenous learners	Calls for systemic shifts and stakeholder collaboration

Table 1 (cont.)

10	Walker et al. (2023)	Provisional principles for culturally responsive assessment ethics	Incorporates student backgrounds, values, and lived experiences	Responds to critiques of cultural bias in large-scale assessments	Principles aim to reduce bias and promote equity	Emphasizes inclusive design and cultural representation
11	Taylor & Evans (2025)	Ethical and social justice perspectives on culturally responsive assessment	Integration of cultural values in large-scale and classroom assessments	Historical grounding in dominant culture critique	Supports equity through culturally relevant grading and feedback	Highlights policy and practice implications for inclusivity
12	Englert & Shultz (2025)	Culture- and worldview-grounded ethical development processes	Emphasizes partnerships with communities for cultural validity	Calls for critical interrogation of assessment foundations	Enhances fairness by aligning assessment with cultural relevance	Advocates for community collaboration and theory of action
13	Taylor & Ferrara (2025)	Ethical critique of psychometric practices suppressing minoritized students	Identifies bias introduced in test development stages	Highlights the dominant culture's influence on measurement	Recommends research and practice changes for equity	Suggests psychometric reform for cultural responsiveness
14	Jankowski & Baker (2025)	Ethical Role of Assessment in culturally responsive curriculum	Supports student engagement and choice in assessment	Links assessment ethics to curriculum inclusivity	Addresses past learning trauma through culturally responsive assessment	Encourages integration of assessment and curriculum design
15	Kerr & Averill (2024)	Ethical design for culturally sustaining assessment of Indigenous learners	Incorporates Māori perspectives and culturally sustaining pedagogy	Policy and professional learning are critical for ethical practice	Enhances learner well-being and maximizes potential	Calls for cultural understanding and supportive policies
16	Rosheim et al. (2024)	Ethical concerns about validity and reliability for multilingual literacy assessment	Highlights cultural and linguistic relevance in assessment materials	Historical issues with deficit-based assessment models	Identifies tensions between language acquisition and assessment scores	Advocates for nuanced selection and awareness in assessment
17	Ghanbari (2024)	Ethical critique of Western-centric assessment frameworks	Calls for indigenized, culturally appropriate assessment models	Historical neglect of minority learners' epistemologies	Links inequitable outcomes to culturally inappropriate frameworks	Proposes culturally grounded frameworks to reduce dropout rates
18	Poudel (2024)	Ethical agency of teachers in decolonizing assessment practices	Highlights constraints of normative standards on cultural responsiveness	Examines the persistence of colonial assessment practices	Identifies epistemic injustice in assessment implementation	Advocates for teacher autonomy and policy reform for equity

Table 1 (cont.)

19	McIntosh (2024)	Autoethnographic ethical critique of deficit ideologies in TESOL	Exposes institutional racism in language assessment	Reflects on historical and ongoing colonial influences	Reveals negative impacts on linguistic minorities	Calls for decolonizing assessment through critical self-reflection
20	Bérešová (2014)	Ethical awareness focused on fairness and washback effects	Examines political impacts on test design and scoring	Tracks the ethical implications of test conditions over time	Ethical design improves fairness and positive washback	Highlights item writers' ethical responsibilities
21	Rasooli & DeLuca (2024)	Theoretical foundations of fairness in classroom assessment	Integrates social justice and equity in assessment theory	Reviews evolving fairness concepts in educational contexts	Fairness is linked to inclusive and equitable classroom practices	Calls for theory-driven research to advance fair assessment
22	Plakans & Lee (2025)	Ethical critique of fairness, justice, and criticality in L2 writing	Advocates for multilingual and socially conscious assessment	Historical review of 25 years of writing assessment ethics	Promotes empowerment and challenges exclusionary norms	Calls for teacher and student voice inclusion in assessment
23	Cervatiuc (2024)	Critical language testing ethical framework addressing social power	Emphasizes multilingualism and diversity in assessment	Challenges traditional psychometric focus on neutrality	Aims to prevent discrimination and marginalization	Promotes ethical educational tools, acknowledging diversity
24	Ahmedova (2025)	Ethical principles in teacher knowledge and test design	Highlights ethical practices for fairness, reliability, and validity	Notes challenges in teacher training and assessment diversity	Ethical adherence improves student engagement and development	Recommends integrated ethical frameworks for assessment quality
25	Inbar-Lourie (2021)	Ethical approach using argument-based validation for fairness	Focuses on justice principles in language assessment evaluation	Historical development of ethical evaluation methods	Ethical validation supports fairness and justice in testing	Encourages rigorous ethical evaluation of assessments
26	Solano-Flores (2022)	Ethical standards for test accommodations for English learners	Emphasizes principled design and evaluation of accommodations	Historical standards guide accommodation practices	Accommodations improve fairness when matched to learner needs	Warns against misuse and stresses methodological rigor
27	Quevedo-Camargo & Pinheiro (2021)	Ethical perspective linking teacher stance and assessment tools	Discusses the historical relativism of ethics in language assessment	Traces the development of ethical codes and guidelines	Ethical compliance enhances classroom assessment validity	Advocates for ethical principles in assessment design and use
28	Boyd & Davies (2002)	Origins and purposes of ethical codes in language testing	Discusses professional responsibility and openness	Historical dev. of the International Language Testing Association (ILTA) Code of Ethics	Ethical codes promote transparency and professional conduct	Notes the challenges of universal codes and local adaptations

Table 1 (cont.)

29	Rodríguez (2004)	Ethical analysis of evaluation impacts and examiner codes	Explores public distrust and ethical commitments in testing	Historical tensions between altruism and coercion in assessment	Ethical measures proposed to restore trust and fairness	Highlights examiner responsibility and ethical conduct
30	Haney & Madaus (1991)	Historical overview of ethical and technical test standards	Documents major revisions in ethical standards over the decades	Tracks the evolution of APA and testing standards	Ethical standards improve test fairness and professional conduct	Suggests ongoing refinement of ethical and technical standards
31	Pandey (2024)	Ethical integration in modern inclusive language testing	Highlights fairness and socio-political dimensions in test design	Reviews shift from traditional to innovative assessment ethics	Ethical practices promote equitable and authentic testing	Addresses technology's role and challenges in ethical assessment
32	Galaczi & Pastorino-Campos (2025)	Ethical principles for AI in language assessment	Advocates an ethical-by-design approach, balancing risks and values	Contextualizes AI ethics within global policy frameworks	Promotes relational accountability and ethical AI practices	Addresses tensions and risk mitigation in AI assessment use
33	Randall et al. (2023)	Justice-oriented antiracist validity framework	Disrupts white supremacist assessment logics	Calls for re-orientation toward assessment justice	Enhances fairness through critical quantitative methods	Encourages antiracist principles in validity evidence gathering
34	Bernard (2023)	Ethical analysis of assessments' impact on marginalized Afro-descendant students	Promotes culturally responsive curriculum and alternative assessments	Historical context of assessment effects on marginalized groups	Ethical adjustments improve educational outcomes and equity	Recommends policy and practice changes for inclusivity
35	Gelvez-Manrique et al. (2025)	Ethical-epistemological critique of situated educational evaluation	Proposes transformative, justice-oriented assessment praxis	Challenges neutral technical views of assessment	Ethical evaluation addresses identity and sociocultural factors	Offers an analytical framework for ethical reflection and agency
36	Ulviani (2025)	Ethical promotion of culturally responsive Indonesian language instruction	Integrates cultural backgrounds into inclusive education	Critiques standardization and exam-driven policies	Ethical practices enhance engagement and social justice	Calls for systemic reforms in policy and teacher education
37	Bhatti & Bashir (2025)	Ethical concerns about cultural integration in English assessment	Examines cultural representation in textbooks and tests	Highlights the imbalance favoring Western perspectives	Cultural bias risks fairness and learner motivation	Suggests culturally balanced content for equitable assessment

Table 1 (cont.)

38	Schissel (2023)	Ethical critique of bias, discrimination, and social consequences	Examines social-political-historical contexts of assessment	Highlights limitations of traditional psychometric paradigms	Calls for socially just assessment practices and validity	Emphasizes critical examination of power and equity in testing
39	Prayitno et al. (2024)	Ethical framework for culturally relevant pedagogy promoting equity	Integrates cultural identities into teaching and assessment	Addresses systemic challenges in culturally relevant pedagogy (CRP) implementation	Ethical CRP enhances motivation, achievement, and social justice	Recommends professional development and institutional support
40	Wibowo et al. (2024)	Ethical considerations in grade inflation among EFL teachers	Explores teacher dilemmas and unconscious biases	Historical pressures influence grading ethics	Ethical grading practices maintain integrity and fairness	Calls for teacher awareness and ethical grading procedures

Table 1 shows a holistic picture of 1) ethical frameworks, 2) cultural responsiveness integration, 3) historical evolution of standards, 4) impact on assessment outcomes, and 5) implementation challenges and solutions. In response to the research questions, these aspects of language assessment are presented as follows:

- 1) Ethical frameworks in language assessment have evolved from traditional ethical codes to more dynamic models emphasizing fairness, accountability, and social justice (e.g., [Boyd & Davies, 2002](#); [Randall et al., 2023](#)). A growing subset of studies integrates culturally responsive and decolonizing ethics that foreground the experiences of marginalized and indigenous communities (e.g., [McIntosh, 2024](#); [Poudel, 2024](#); [Steele et al., 2024](#)). Ethical validation has been tied to professionalization, stakeholder collaboration, and responsible test use (e.g., [Bachman, 2024](#); [Solano-Flores, 2022](#)), while emergent discussions stress the incorporation of AI ethics and relational accountability in assessment design (e.g., [Galaczi & Pastorino-Campos, 2025](#); [Pandey, 2024](#)).
- 2) Cultural responsiveness is increasingly recognized as central to ethical assessment. Most studies underscore the need to embed learners' cultural and linguistic diversity in assessment design and implementation (e.g., [Solano-Flores, 2019](#); [Walker et al., 2023](#)). It is also essential to consider pedagogical approaches that validate students' identities and lived experiences (e.g., [Baidoo-Anu et al., 2023](#); [Jankowski & Baker, 2025](#); [Prayitno et al., 2024](#)). However, critiques of assessment and standardized testing tend to emphasize persistent sociocultural biases, advocating for inclusive, culturally sustaining models ([Bhatti & Bashir, 2025](#); [Rosheim et al., 2024](#)). To minimize bias, collaborative and participatory practices involving communities and stakeholders are deemed essential strategies to strengthen cultural responsiveness ([Englert & Shultz, 2025](#); [Poudel, 2024](#)).
- 3) Historical evolution of standards reflects a gradual transformation from early ethical codes toward frameworks that promote equity and social justice (e.g., [Boyd & Davies, 2002](#); [Haney & Madaus, 1991](#); [Shohamy, 1997](#)). However, remnants of colonial and Western-centric norms still exist, prompting calls for decolonization and inclusive reform ([Ma, 2025](#); [McIntosh, 2024](#)). Historical critiques of standardized testing have spurred the development of culturally responsive assessment models ([Brunfaut, 2023](#); [Schissel,](#)

- 2023), while contemporary standards increasingly engage with AI-driven ethical considerations (Burstein & LaFlair, 2024; Galaczi & Pastorino-Campos, 2025).
- 4) Impact on assessment outcomes is a concerning issue. Empirical evidence emphasizes strong ethical frameworks that further enhance fairness, validity, and equity in language assessment outcomes (e.g., Ma, 2025; Randall et al., 2023). Culturally responsive approaches could also reduce construct-irrelevant variance and mitigate bias (e.g., Kerr & Averill, 2024; Solano-Flores, 2019; Walker et al., 2023), leading to positive washback effects and more equitable educational opportunities (Bérešová, 2014; Bernard, 2023; Wibowo et al., 2024). Nonetheless, systemic and sociopolitical inequities continue to challenge the realization of full fairness (Schissel, 2023).
 - 5) Implementation challenges and solutions are well-informed. Despite significant progress, challenges—such as standardization pressures, inadequate training, and institutional constraints—hinder ethical implementation (Baidoo-Anu et al., 2023; Ghanbari, 2024; Prayitno et al., 2024). However, solutions are subject to focus on participatory design, stakeholder engagement, and culturally grounded frameworks (Englert & Shultz, 2025; Poudel, 2024; Rasooli & DeLuca, 2024). Similarly, previous studies (e.g., Prayitno et al., 2024; Solano-Flores, 2022; Wibowo et al., 2024) stress the importance of teacher education, policy reform, and ethical vigilance in areas such as high-stakes grading and accommodations. Despite technologically integrated opportunities for innovative assessment and testing, it is also possible to raise new ethical challenges (Pandey, 2024; Burstein & LaFlair, 2024), especially systemic inequities (Gelvez-Manrique et al., 2025; Schissel, 2023).

In a nutshell, ethics and language assessment standards are a growing recognition of the complex interplay between ethical principles, cultural responsiveness, and the historical evolution of assessment practices. A prominent theme is the call for justice-oriented and culturally sustaining approaches that challenge traditional, often Western-centric, paradigms. While several studies provide robust theoretical frameworks and practical recommendations for enhancing fairness and inclusivity, there remain significant gaps in empirical validation and implementation, particularly in diverse sociocultural contexts. Moreover, the integration of emerging technologies such as AI introduces new ethical considerations. Overall, the body of research underscores the necessity of ongoing critical reflection and collaborative efforts among stakeholders to realize equitable language assessment.

Language Assessment in Sociocultural-Historical Contexts

Based on the thematic analysis of the reviewed studies ($n = 40$), key themes were found to focus on fairness, cultural responsiveness, and the evolution of ethical frameworks in assessment practices. However, predominant discussions tend to focus on how cultural and social justice considerations shape test design, validation, and implementation across diverse learner populations. Historical developments and emerging technologies could further inform contemporary ethical debates, with a strong emphasis on dismantling systemic biases and promoting inclusivity. The following is a thematic summary of ethics and language assessment standards in sociocultural-historical contexts, as shown in Table 2.

Table 2
Key Themes of the Reviewed Studies (n = 40)

Key Theme	Description
Fairness, Justice, and Ethical Standards in Language Assessment (e.g., Bérešová, 2014; Haney & Madaus, 1991; Plakans & Lee, 2025; Rasooli & DeLuca, 2024; Schissel, 2023; Solano-Flores, 2022)	Fairness and justice are the core pillars of ethical language assessment, emphasizing equitable access, validity, and the mitigation of bias throughout test development, administration, and use. Research highlights the necessity of ethical codes, quality control mechanisms, and stakeholder collaboration to uphold these standards, often invoking frameworks such as Messick's (1989) validity and Kane's (2010) fairness principles. The literature underscores that ethical standards profoundly impact assessment outcomes and social equity across diverse contexts.
Cultural Responsiveness and Inclusivity in Assessment (e.g., Baidoo-Anu et al., 2023; Kerr & Averill, 2024; Rosheim et al., 2024; Solano-Flores, 2019; Taylor & Evans, 2025; Taylor & Ferrara, 2025; Ulviani, 2025; Walker et al., 2023)	Cultural responsiveness is extensively recognized as essential for valid and fair language assessment. Studies advocate for assessment models and practices that reflect students' cultural identities, experiences, and linguistic diversity, addressing cultural bias and promoting inclusivity. Emphasis is placed on collaborative design, cultural validity, and the integration of culturally sustaining pedagogies to improve learner engagement and social justice.
Historical and Theoretical Evolution of Ethical Language Assessment (e.g., Bachman, 2024; Boyd & Davies, 2002; Haney & Madaus, 1991; Pandey, 2024; Rodríguez, 2004; Shohamy, 1997; Wang, 2024)	The historical trajectory of ethical standards in language assessment is traced through revisions of professional codes, validation theories, and shifts from traditional to modern paradigms that incorporate ethics centrally. Seminal works document how ethical awareness evolved alongside technical advancements and the professionalization of language testing, highlighting ongoing challenges in operationalizing ethics across contexts.
Impact of Ethics on Assessment Implementation and Outcomes (e.g., Bérešová, 2014; Liu et al., 2024; Ma, 2025; Schissel, 2023)	Ethical considerations significantly shape assessment design, implementation, and interpretation, influencing validity, fairness, and learner experiences. Empirical studies reveal how ethical lapses can reinforce systemic inequalities, while ethical frameworks guide reforms enhancing assessment validity and equity. Integration of ethical standards is shown to improve the meaningfulness of scores and reduce adverse consequences on marginalized populations.
Decolonization and Anti-Racist Approaches in Language Testing (e.g., Englert & Shultz, 2025; McIntosh, 2024; Poudel, 2024; Randall et al., 2023)	A growing body of research advocates for decolonizing language assessments by confronting colonial legacies, white supremacy, and epistemic injustices embedded in testing practices. Justice-oriented and antiracist validation frameworks aim to disrupt oppressive systems, promote linguistic and cultural legitimacy, and empower marginalized learners. This emergent theme emphasizes community partnerships and critical reflection on assessment purposes and foundations.
Ethical Challenges and Professional Responsibility in Classroom and Large-scale Assessment (e.g., Bérešová, 2014; Liu et al., 2024; Solano-Flores, 2022; Wibowo et al., 2024)	Language assessment ethics extend to classroom and large-scale contexts, where professional responsibility includes addressing grade inflation, test bias, accommodations, and transparency. Studies explore teachers' ethical judgments, quality control practices, and the role of codes of ethics in guiding responsible assessment to foster fairness and reliability. The literature identifies tensions between policy demands and ethical implementation.
Technology and Ethical Considerations in Language Assessment (e.g., Burstein & LaFlair, 2024; Galaczi & Pastorino-Campos, 2025; Pandey, 2024)	Advances in technology, including AI and computer-adaptive testing, introduce new ethical considerations related to fairness, validity, and social responsibility. Frameworks integrating classical validation theory with responsible AI principles support ethical use of automated scoring and assessment technologies. The literature calls for "ethical-by-design" approaches to mitigate risks and uphold human values.

Key Theme	Description
Cross-Cultural Variation and Ethical Judgment in Assessment Practices (e.g., Bernard, 2023; Liu et al., 2024; Schissel, 2023)	Cross-cultural studies reveal variations in ethical judgments and assessment practices, underscoring the influence of sociocultural contexts on fairness perceptions and ethical decision-making. Comparative research involving diverse educational settings highlights the need for culturally sensitive ethical frameworks that acknowledge local values and traditions.
Emerging Frameworks for Culturally Sustaining and Responsive Assessment (e.g., Ghanbari, 2024; Kerr & Averill, 2024; Prayitno et al., 2024; Taylor & Ferrara, 2025; Ulviani, 2025)	Recent contributions propose provisional principles and measurement arguments for culturally responsive and sustaining assessments that go beyond inclusion to actively validate diverse learner identities and epistemologies. These frameworks emphasize proactive design and ongoing evaluation to align assessments with sociocultural realities and promote equity.

The reviewed studies ($n = 40$) are based on ethics and language assessment standards that are operationalized in sociocultural-historical contexts. The results of the analysis show key themes that situate fairness, justice, and cultural responsiveness at the heart of assessment practices. Additionally, there is a mutual agreement that ethical considerations should extend beyond procedural validity to encompass the social and political contexts in which assessments operate. Some scholars like Randall et al. (2023) and Schissel (2023) emphasize the necessity of disrupting white supremacist legacies and reimagining assessment as a socially situated act of justice. These perspectives are consistent with Bachman (2024), stressing the tie between ethical conduct and test validity. Besides, cultural responsiveness has emerged as a dominant theme, with several studies (e.g., Ojochegbe, 2024; Steele et al., 2024) advocating for assessment frameworks that reflect learners' identities, linguistic diversity, and sociocultural realities.

However, despite theoretical sophistication, implementation remains uneven. Institutional inertia, lack of professional training, and pressures of standardization seem to hinder progress. As noted by Baidoo-Anu et al. (2023) and Poudel (2024), effective realization of ethical and culturally responsive assessment requires systemic support, participatory design, and teacher agency. Another point of discussion is the persistent tension between psychometric rigor and cultural relevance, despite advancements in validation theory—particularly argument-based and justice-oriented validation—offering directions. In reality, the operationalization of such frameworks remains limited. Furthermore, the emergence of AI in assessment introduces novel ethical dilemmas concerning transparency, accountability, and algorithmic bias (Burstein & LaFlair, 2024; Galaczi & Pastorino-Campos, 2025). This technological frontier requires attention to ensure that innovations enhance rather than undermine fairness and inclusivity. From a sociopolitical standpoint, the reviewed studies reveal that ethical inattention in assessment is not merely a technical oversight but a reflection of broader inequities. Studies on decolonizing and antiracist assessment (e.g., Englert & Shultz, 2025; McIntosh, 2024) argue that addressing these inequities demands critical reflexivity, community collaboration, and institutional reform. These approaches reframe ethics not as compliance with universal codes, but as context-sensitive praxis grounded in social justice and cultural sustainability.

Besides, one critical gap lies in limited attention to emerging technologies, particularly artificial intelligence (AI). Although recent work acknowledges the ethical challenges of automated assessment systems, studies such as Burstein & LaFlair (2024) and Galaczi & Pastorino-Campos (2025) indicate that responsible AI integration in testing remains limited, with few empirical validations or consensus-driven guidelines. Similarly, the focus on

standardized testing has overshadowed ethical issues in classroom-based and formative assessments, which play an equally crucial role in shaping learner experiences. Persistent cultural bias also undermines the inclusivity and fairness of language assessment practices. Despite increased recognition of cultural responsiveness, most studies (e.g., Ojohegbe, 2024; Walker et al., 2023; Saragih, 2024) show that integrating cultural diversity into test design and implementation remains inconsistent, hindered by institutional rigidity, lack of professional development, and limited policy support. Moreover, historical and sociopolitical dimensions of assessment ethics are often underemphasized, leaving the systemic legacies of colonialism and inequality insufficiently addressed.

CONCLUSION

In conclusion, the reviewed studies depict an evolving connection between ethics, culture, history, and practice in language assessment. Ethical frameworks have shifted beyond fairness and validity toward justice-oriented, antiracist, and decolonizing perspectives that challenge Western-centric norms and systemic inequities. These frameworks highlight transparency, professional responsibility, and social justice, positioning assessment as a socially situated act that can either perpetuate or disrupt oppression. Cultural responsiveness remains central, with an emphasis on the integration of learners' diverse backgrounds and experiences to enhance validity and fairness while minimizing cultural bias. However, these often remain conceptual as a consequence of standardization pressures, limited training, and institutional resistance. Collaborative and participatory approaches involving stakeholders and communities show promise but need broader implementation. Historical and sociopolitical analyses expose enduring colonial and power structures in assessment, with critical and decolonizing methodologies offering pathways for reform—though practical tools for large-scale change remain scarce. Finally, the rise of AI introduces new ethical challenges that need to be reconciled with traditional principles of validity, fairness, and accountability to ensure assessments serve equity and justice across contexts.

Looking forward, future research should prioritize empirical validation of justice-oriented and culturally responsive frameworks. This may include testing antiracist and decolonizing models across varied contexts to assess their scalability and real-world impact. There is also a need to bridge the gap between ethical codes and practice, ensuring that professional guidelines translate into actionable standards in both classroom and large-scale testing environments. Studies should investigate the barriers preventing ethical adherence and propose mechanisms for accountability and capacity building. Additionally, researchers are encouraged to explore teacher agency in decolonizing assessment, empowering educators to adapt practices that reflect local epistemologies and learner identities. The development of innovative psychometric methods that incorporate cultural validity and epistemic pluralism is essential to reconcile technical rigor with social equity. Expanding professional development programs for teachers and test designers remains vital to sustain ethical and inclusive assessment cultures. Finally, as AI continues to transform educational assessment, future studies should focus on establishing ethical-by-design frameworks ensuring transparency, fairness, and human oversight in automated testing. Addressing these directions will not only narrow existing research gaps but also pave the way for language assessment systems that are empirically grounded, culturally responsive, and ethically sound across global contexts.

REFERENCES

- Ahmedova, M. (2025). Scaling of students' language skills in English as a foreign language. *EuroGlobal Journal of Linguistics and Language Education*, 2(1), 24–32. <https://doi.org/10.69760/egjille.250003>
- Bachman, L. F. (2024). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing*, 17(1), 1–42. <https://doi.org/10.1177/026553220001700101>
- Baidoo-Anu, D., Lei, L., Cisterna, D., & Song, Y. (2023). Cultural validity: Promoting cultural responsiveness in classroom assessment. *Diaspora, Indigenous, and Minority Education*, 18(3), 160–181. <https://doi.org/10.1080/15595692.2023.2193883>
- Bérešová, J. (2014). Ethical awareness and fairness in language testing. *European Scientific Journal*, 10(10), 572–577. <https://www.academia.edu/120445004>
- Bernard, D. M. (2023). Beyond assessments: Exploring strategies for Afro-Descendant students' success in Costa Rica. *Journal of Social, Humanity, and Education*, 3(3), 253–269. <https://doi.org/10.35912/jshe.v3i3.1475>
- Bhatti, A. M., & Bashir, A. (2025). From Textbooks to Tests: The Role of Cultural Integration in Shaping Outcomes of English Language Assessment in Punjab, Pakistan. *Research Square*. <https://doi.org/10.21203/rs.3.rs-7583969/v1>
- Boyd, K., & Davies, A. (2002). Doctors' orders for language testers: the origin and purpose of ethical codes. *Language Testing*, 19(3), 296–322. <https://doi.org/10.1191/0265532202lt231oa>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brignardello-Petersen, R., & Guyatt, G. H. (2025). Assessing the certainty of the evidence in systematic reviews: importance, process, and use. *American journal of epidemiology*, 194(6), 1681–1686. <https://doi.org/10.1093/aje/kwae332>
- Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.
- Burstein, J., & LaFlair, G. T. (2024). Where assessment validation and responsible AI meet. *Language Teaching Research Quarterly, Special Issue*, 1–31. <https://doi.org/10.48550/arxiv.2411.02577>
- Cervatiuc, A. (2024). The social dimension of language testing. In J. I. Lontas (Ed.), *The TESOL encyclopedia of language teaching* (pp. 1–7). Wiley.
- Englert, K., & Shultz, P. K. (2025). Reconsidering theories of action. In C. M. Evans & C. S. Taylor (Eds), *Culturally Responsive Assessment in Classrooms and Large-Scale Contexts: Theory, Research, and Practice* (pp. 156–176). Routledge.
- Galaczi, E., & Pastorino-Campos, C. (2025). Ethical AI for language assessment: Principles, considerations, and emerging tensions. *Annual Review of Applied Linguistics*, 45, 294–314. <https://doi.org/10.1017/S0267190525100081>
- Gelvez-Manrique, J. D., Parada-Carreño, M. J., & Hernández-Albarracín, J. D. (2025). Beyond technicism: ethical-epistemological foundations for a situated and critical educational assessment. *AiBi Journal of Research, Administration and Engineering*, 13(2), 1–8. <https://doi.org/10.15649/2346030X.5422>
- Ghanbari, H. (2024). Educational Reform Through Designing Culturally Appropriate Assessment Frameworks. *Journal of Comparative & International Higher Education*, 16(4), 45–51. <https://doi.org/10.32674/y2mw3n24>

- Haney, W., & Madaus, G. F. (1991). The Evolution of Ethical and Technical Standards for Testing. In R. K. Hambleton & J. N. Zaal (Eds), *Advances in Educational and Psychological Testing: Theory and Applications* (pp. 395–425). Springer Dordrecht.
- Inbar-Lourie, O. (2021). Evaluating language assessments. *Language Assessment Quarterly*, 18(2), 199–203. <https://doi.org/10.1080/15434303.2021.1872577>
- Jankowski, N. A., & Baker, G. R. (2025). The role of culturally responsive assessment in culturally responsive curriculum. In R. Rantz & L. McNulty (Eds), *Developing Culturally Responsive Curriculum in Higher Education* (pp. 439–452). IGI Global.
- Jesson, J., Matheson, L., and Lacey, F. (2011). *Doing Your Literature Review: Traditional and Systematic Techniques*. Sage Publishing.
- Kerr, B. G., & Averill, R. (2024). Culturally sustaining assessment for indigenous learners in aotearoa New Zealand: Drawing from policy, research, and educator voices. *Teaching and Teacher Education*, 143(1), 1–10. <https://doi.org/10.1016/j.tate.2024.104567>
- Liu, J., Bon, S., Fan, X., Hardie, S., & Ding, R. (2024). A comparative mixed-methods study of in-service teachers' ethical judgment about student assessment practices. *Psychology in the Schools*, 61(9), 3441–3464. <https://doi.org/10.1002/pits.23226>
- Ma, Z. (2025). From Validity to Equity: Revisiting CET-4 Through Messick, Kane, and Paraskeva's Frameworks. *International Journal of Language and Linguistics*, 13(3), 147–157. <https://doi.org/10.11648/j.ijll.20251303.14>
- McIntosh, L. (2024). Language and Decolonisation: An Interdisciplinary Approach. In F. Ndhlovu & S. J. Ndhlovu-Gatsheni (Eds), *Decolonising competency-based assessment through autoethnographic approaches*. Routledge.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, 151, 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>
- Randall, J., Poe, M., Oliveri, M. E., & Slomp, D. (2023). Justice-Oriented, Antiracist Validation: Continuing to Disrupt White Supremacy in Assessment Practices. *Educational Assessment*, 29(1), 1–20. <https://doi.org/10.1080/10627197.2023.2285047>
- Schissel, J. L. (2023). Bias, Discrimination, and the Social Consequences of Unproblematized Assessments in TESOL. *TESOL Quarterly*, 57(2), 716–721. <https://doi.org/10.1002/tesq.3219>
- Shohamy, E. (1997). Testing methods, testing consequences: are they ethical? Are they fair?. *Language Testing*, 14(3), 340–349. <https://doi.org/10.1177/026553229701400310>
- Solano-Flores, G. (2019). Examining Cultural Responsiveness in Large-Scale Assessment: The Matrix of Evidence for Validity Argumentation. *Frontiers in Education*, 4, 1–9. <https://doi.org/10.3389/educ.2019.00043>
- Solano-Flores, G. (2022). Fairness in Testing: Designing, Using, and Evaluating Test Accommodations for English Learners. In J. L. Jonson & K. F. Geisinger (Eds), *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standards* (pp. 271–292). American Educational Research Association.

- Steele, C., Gower, G., & Bogachenko, T. (2024). Creating and enacting culturally responsive assessments for First Nations students in higher education settings. *Australian Journal of Education*, 68(2), 84–102. <https://doi.org/10.1177/00049441241258496>
- Taylor, C. D., & Evans, C. M. (2025). *Culturally responsive assessment*. Routledge. https://api.pageplace.de/preview/DT0400.9781040337448_A50972424/preview-9781040337448_A50972424.pdf
- Taylor, C. D., & Ferrara, S. (2025). A measurement argument for culturally responsive assessment. In C. M. Evans & C. S. Taylor (Eds), *Culturally Responsive Assessment in Classrooms and Large-Scale Contexts: Theory, Research, and Practice* (pp. 135–155). Routledge.
- Ojochegbe, A. T. (2024). Rethinking Standardized Testing in English Language Proficiency: Moving Toward Culturally Responsive Assessment Models. *Jurnal Pendidikan Indonesia*, 5(12), 1990–1996. <https://doi.org/10.59141/japendi.v5i12.6584>
- Pandey, G. P. (2024). From traditional to modern: Evolving practices in language testing and assessment. *Interdisciplinary research in education*, 9(2), 26–37. <https://doi.org/10.3126/ire.v9i2.75020>
- Plakans, L., & Lee, K. (2026). Fairness, justice, and criticality: Reviewing second language writing assessment. *Language Teaching*, 59(1), 1–28. <https://doi.org/10.1017/s0261444825100876>
- Poudel, P. P. (2024). Decolonization of English language assessment for equitable learning: Teachers' agency perspective. *Journal of Nepalese Studies*, 16(1), 140–152. <https://doi.org/10.3126/jns.v16i1.71787>
- Prayitno, H., Cahyo, D. H., Rasinus, R., Lismina, L., & Rosita, R. (2024). Culturally relevant pedagogy: Promoting inclusivity and equity in the classroom. *Global international journal of innovative research*, 2(3), 728–735. <https://doi.org/10.59613/global.v2i3.118>
- Quevedo-Camargo, G., & Pinheiro, L. A. (2021). Ethics in additional language assessment: from the teacher's attitude to the assessment instrument [Translated version]. *Estudos em Avaliação Educacional*, 32, e08641. <https://doi.org/10.18222/eae.v32.8641>
- Rasooli, A. H., & DeLuca, C. (2024). A critical review of fairness from multiple perspectives: Implications for classroom assessment theory. *Applied Measurement in Education*, 37(2), 148–164. <https://doi.org/10.1080/08957347.2024.2345594>
- Rodríguez, J. R. P. (2004). Ethical aspects of assessment: the impact of evaluative activities and the deontological codes of examiners [Translated version]. *Carabela*, 55, 31–43. https://cvc.cervantes.es/ensenanza/biblioteca_ele/carabela/pdf/55/55_031.pdf
- Rosheim, K. M., Tamte, K. G., & Froemming, M. J. (2024). Reducing inequalities inherent in literacy assessment of multilingual learners. *Reading Psychology*, 45(8), 735–751. <https://doi.org/10.1080/02702711.2024.2359922>
- Saragih, E. (2024). EFL lecturers' practice, perception, and evaluation of English as the medium of instruction in tertiary education. *Issues in language studies*, 13(1), 93–111. <https://doi.org/10.33736/ils.5789.2024>
- Ulviani, M. (2025). Culturally Responsive Approaches to Indonesian Language Instruction: Insights for Inclusive Education. *Research Square*. <https://doi.org/10.21203/rs.3.rs-7526656/v1>

Chanpradit & Saragih (2026): *Ethics and Language Assessment Standards...*

Walker, M. E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D., & Gholson, M. (2023). *Culturally responsive assessment: Provisional principles* (Research Report No. RR-23-11). ETS. <https://doi.org/10.1002/ets2.12374>

Wang, J. (2024). Ethics and context in second language testing: rethinking validity in theory and practice. M. Rafael Salaberry, Albert Weideman, and Wei-Li Hsu (2023). New York, Routledge, 248 Pages, ISBN 9781032471754. *Porta Linguarum*, 42, 359–362. <https://doi.org/10.30827/portalin.vi42.30001>

Wibowo, E. P., Lestari, A., Adrian, M., Firdausiyah, K. S., Islamiy, J. H., Putri, R. D., Kemal, T., & Kurniawan, E. (2024). Exploring the Ethical Dimensions of Testing and Assessment: An Investigation into Grade Inflation Among EFL Teachers. *OKARA: Jurnal Bahasa Dan Sastra*, 18(1), 71–97. <https://doi.org/10.19105/ojbs.v18i1.12676>