Available online at: https://jurnal-fkip.ut.ac.id/index.php/ijrse/issue/archive
International Journal of Research in STEM Education (IJRSE)
ISSN 2721-2904 (online)
Volume 8 Number 1 (2026): 11 - 29

Equity-Oriented Formative Assessment in K–12 STEM Classrooms: *A Systematic Review*

[1]Bilgehan Ayik, [2]Dai Gu, [3]Woomee L. Kim, [4]Marjorie H. Haley
[1]College of Education and Human Development, George Mason University, USA
[1]bayik.gmu.edu

Abstract

Multilingual learners (MLs) remain underrepresented in science, technology, engineering, and mathematics (STEM) related careers and minority in K-12 STEM education, partly due to inequitable assessment practices that fail to account for their cultural and linguistic diversity. Culturally and linguistically responsive (CLR) formative assessment has the potential to enhance equity and inclusivity in K-12 STEM classrooms. This systematic literature review examines the key trends and challenges in CLR formative assessment practices for MLs in STEM education in the United States. Through the systematic search and elimination process, 25 articles (12 empirical and 13 non-empirical) were selected, analyzed and discussed. Findings highlighted key trends, including collaborative design, translanguaging, multimodal assessment, and tailored feedback, while also identifying persistent challenges such as balancing of STEM content learning goals with language acquisition, professional development (PD) gaps, and the marginalization of MLs through standardized language policies in mainstream K-12 STEM classrooms. Implications highlighted the need for formative assessment frameworks that center equity, policy reforms to support linguistically inclusive assessment, and PD opportunities that equip educators with CLR formative assessment strategies.

Keywords: *Culturally and linguistically responsive assessment, Formative assessment, Multilingual learners, STEM, Translanguaging*

INTRODUCTION

The U.S. is witnessing a noticeable increase in the K-12 student population of culturally and linguistically diverse students (Ramírez & Faltis, 2020; Ramírez, 2023). As of Fall 2021, 10.6% of students in U.S. public schools were classified as English Learners, or, using a more asset-oriented term, Multilingual Learners (MLs) (National Center for Education Statistics [NCES], 2024). MLs have been notably underrepresented in majors related to science, technology, engineering, and mathematics (STEM) within higher education and the professional fields (Stephens & Francis, 2018). This underrepresentation is closely associated with the achievement gap between historically minoritized ethnic and racial groups in STEM subjects (White House Office of Science and Technology Policy, 2023). Formative assessments have the potential to bridge achievement gaps - or what Ladson-Billings (2006) reframes as education debts - by offering insights that improve teaching for all students (Furtak et al., 2016). Therefore, focusing on formative assessments in multicultural and multilingual classrooms could be a key strategy for addressing these disparities.

As classrooms become increasingly multilingual and culturally diverse, formative assessment for MLs requires educators to attend to students' linguistic and cultural resources. When assessments are not responsive to students' linguistic and cultural backgrounds, they may provide an incomplete picture of what MLs know and can do, potentially reinforcing existing inequities in STEM education (Ayik et al., 2026; Wolf et al., 2025). Consequently, scholars have called for culturally and linguistically responsive (CLR) approaches to assessment

Corresponding author:
Bilgehan Ayik, bayik@gmu.edu
DOI: https://doi.org/10.33830/ijrse.v8i1.14021

that recognize and build upon students' assets rather than viewing linguistic diversity through a deficit lens (Ntumi et al., 2026; Robinson-Jones et al., 2026). Recent studies have further emphasized the value of translanguaging and other culturally sustaining practices that allow MLs to draw on their full linguistic repertoires while engaging in STEM learning and assessment (Cheung & Ng, 2026; Kim et al., 2026). In addition, emerging research suggests that language development, disciplinary learning, and assessment are deeply interconnected processes, particularly in multilingual STEM classrooms, and should be understood as such (Aguirre-Muñoz et al., 2024; Lyon, 2022). This perspective aligns with sociocultural views of formative assessment, which emphasize the role of classroom interactions and instructional responses in supporting student learning (Heritage, 2010; Trumbull & Lash, 2013). Despite growing attention to equity, multilingualism, and STEM participation, relatively little research has examined how these ideas come together within formative assessment practices, and CLR assessment remains an underexplored area in U.S. schools (Ayik et al., 2026; Olivera-Aguilar et al., 2024; Yan, 2024).

Although this review focuses on the U.S. context, the identified patterns and challenges reflect broader issues in global STEM education, particularly in contexts characterized by increasing cultural and linguistic diversity. As such, the findings may offer transferable insights for international efforts to design more equitable and inclusive formative assessment practices. This study contributes to the literature by providing a systematic synthesis that integrates formative assessment theory, sociocultural perspectives, and equity-oriented pedagogies within STEM education for MLs. In doing so, it offers a conceptual lens for understanding how assessment practices shape opportunities for participation and learning in culturally and linguistically diverse classrooms.

Research Objectives:

To address the identified gap in the literature, this study conducted a systematic literature review examining CLR formative assessment practices for MLs in K-12 STEM education in the United States. Specifically, this review was designed to:

1. examine how existing formative assessment practices have been implemented in multilingual and multicultural STEM classrooms;
2. explore how issues of equity, inclusion, and linguistic diversity have been addressed within formative assessment practices for MLs;
3. identify common approaches and strategies used to support CLR assessment in STEM education;
4. investigate challenges and barriers associated with implementing CLR formative assessment practices in K–12 STEM settings;
5. discuss how formative assessment practices shape opportunities for MLs' participation, engagement, and meaning-making in STEM learning environments; and
6. propose strategies and implications for implementing more equitable and CLR formative assessment practices in diverse STEM classrooms.

Guided by these objectives, this purposeful review addressed the following research question: "What are the key trends and challenges in CLR formative assessment practices for MLs in K-12 STEM education, with a focus on equity and inclusivity?"

LITERATURE REVIEW

Formative Assessment, Sociocultural Theory, and Equity in K–12 STEM Classrooms

Formative assessment refers to instructional practices through which teachers gather evidence of students' learning in order to guide subsequent teaching and learning. The term was initially coined by Scriven (1967) and further elaborated by Bloom and colleagues (Bloom et al., 1971; Newton, 2007). Early conceptualizations emphasized identifying students' developing understandings and using this information to support progress toward learning goals rather than assign grades or certify achievement (Bloom et al., 1971; Newton, 2007). In this sense, formative assessment is fundamentally oriented toward supporting learning processes rather than evaluating completed performance.

Although formative assessment has not historically been tied to a single learning theory (Wiliam, 2010, as cited in Trumbull & Lash, 2013), many scholars have noted its strong alignment with sociocultural perspectives on learning (Heritage, 2010). Sociocultural constructivist theory perceives learners who actively build and regulate their own knowledge based on their prior knowledge within a social and cultural context (Bruner, 1985; Heritage, 2010; Piaget, 1954; Trumbull & Lash, 2013; Vygotsky, 1978). Within this perspective, formative assessment can be viewed as a process through which teachers elicit evidence of student thinking, interpret that evidence, and respond instructionally to support further learning. Rather than functioning solely as a technique for monitoring progress, formative assessment becomes part of an ongoing interactional process through which teaching and learning are jointly shaped.

Sociocultural framing also has important implications for equity in education. If learning is mediated by social and cultural tools, then formative assessment practices cannot be equitable when they ignore the cultural and linguistic resources students bring to the classroom. Several culture-based pedagogies align closely with the sociocultural constructivist basis of formative assessment. Culture-based pedagogies, as emphasized in the recent report, the Equity in K-12 STEM Education (National Academies of Sciences, Engineering, and Medicine [NASEM], 2024), posit that "culture is everywhere, and all students live and learn culturally, not just certain groups of students" (p. 232). Equipping teachers with culture-based pedagogical approaches that nurture equity-oriented teaching is crucial to improving noticing for equity of teachers in diverse classroom settings (Louie, 2018; Williams et al., 2020).

These culture-based approaches share an asset-oriented orientation toward students' knowledge, experiences, and linguistic repertoire. Although they differ in historical origins and emphasis, collectively they highlight the importance of instructional practices that recognize and build on learners' cultural and linguistic resources. For example, funds of knowledge, one of the prominent culture-based pedagogies, was coined by Vélez-lbáñez (1988) to refer to the knowledge and skills that have been built over time and culturally shaped, which are crucial for individual and household functioning and well-being. Culturally relevant pedagogy was introduced by Ladson-Billings (1995), and it centers the relations of learners' everyday life with the curriculum. Ladson-Billings (1995) explained the criteria of culturally relevant pedagogy as "(a) students must experience academic success; (b) students must develop and maintain cultural competence; and (c) students must develop a critical consciousness through which they challenge the status quo of the current social order" (p. 160). Culturally responsive teaching, developed by Geneva Gay (2018, 2021), centers multiculturalism in the classroom by embracing the assets of students from different ethnic and racial backgrounds and by rejecting the deficit-oriented approaches to underrepresented students. Culturally sustaining and culturally resurgent pedagogies are more recent pedagogies. Culturally sustaining pedagogy underlines the continuity of multiethnic and multilingual learning environments as the key goal of education (Paris, 2012), while culturally resurgent, also known as culturally revitalizing pedagogy, emphasizes resisting colonial impacts, especially within indigenous communities, by supporting the recovery of cultural, spiritual, and linguistic practices and promoting educational sovereignty (Corntassel & Hardbarger, 2019; NASEM, 2024). Rather than treating these

pedagogies as separate add-ons, they can be understood collectively as reinforcing the need for assessment practices that recognize learners' full humanity and support meaningful participation in disciplinary activity.

Bailey et al. (2008) emphasize the crucial role of language by stating that "Classrooms are first and foremost, language environments" (p. 608). However, content educators often lack the training to view the classroom from a linguistic perspective (Trumbull & Lash, 2013). In other words, content teachers, especially those teaching subjects such as mathematics, science, and history, may fail to recognize the need to provide language support to MLs. They also may struggle to effectively support their simultaneous development of content knowledge and language skills in multilingual classrooms. To address this issue, translanguaging pedagogy was developed by scholars, particularly in recent years, to enhance teachers' awareness and abilities to leverage their MLs' full linguistic repertoire to support their diverse learning needs. Translanguaging refers to "both the complex and fluid language practices of bilinguals, as well as the pedagogical approaches that leverage those practices" (García & Lin, 2017, p. 118), which has the potential to nurture equity-oriented instruction and assessment in linguistically diverse classrooms.

These asset-oriented perspectives are especially consequential in STEM education, where participation has historically been shaped by exclusionary assumptions about who belongs, whose knowledge counts, and what forms of language and reasoning are legitimate. Recent calls for equity in K–12 STEM education have underscored that culture is not peripheral to learning but central to how students interpret, engage with, and demonstrate understanding in classrooms (NASEM, 2024). For MLs, the relationship between equity and assessment is particularly important because STEM learning often involves simultaneous engagement with new concepts, specialized discourse practices, and dominant language expectations (Lee & Januszyk, 2019; Llosa et al., 2022). When assessment practices are shaped by monolingual or monocultural assumptions, they may underestimate what MLs know by conflating emergent language practices with limited conceptual understanding. On the other hand, in STEM classrooms, a translanguaging orientation can expand what counts as valid evidence of student understanding by allowing students to draw on their full linguistic repertoires as they explain, question, model, and reason (Kim et al., 2026). When paired with formative assessment, this orientation can help teachers elicit richer evidence of students' developing ideas and make more valid instructional interpretations of what students know and can do.

Foreseeing growing linguistic and cultural diversity, the Next Generation Science Standards emphasized the importance of STEM education for all students, including MLs, as reflected in Appendix D (Lee & Januszyk, 2019; Llosa et al., 2022; National Research Council, 2012). Since then, researchers have increasingly examined the participation and success of diverse student groups in STEM learning environments (Buxton & Lee, 2014; Suh et al., 2020). Previous reviews have also synthesized research on MLs' participation in STEM learning and the role of language practices in science classrooms (e.g., Lee & Januszyk, 2019; Llosa et al., 2022). More recently, NASEM (2024) highlighted broader issues of access and equity across K–12 STEM education systems. In parallel, formative assessment has been examined in influential early reviews that highlight both its promise for supporting classroom learning (Black & Wiliam, 1998) and unresolved conceptual, methodological, and domain-specific concerns in the field (Bennett, 2011). However, these bodies of literature have largely developed separately, leaving limited synthesis on how formative assessment functions specifically within STEM classrooms serving MLs or how assessment practices shape opportunities for these students to demonstrate and develop disciplinary understanding.

The Study Context

Taken together, formative assessment, sociocultural theory, and equity-oriented pedagogies converge in a shared emphasis on learning as interactional, mediated, and shaped by access to meaningful participation. This convergence is particularly important in K–12 STEM classrooms serving MLs, where opportunities to learn are influenced by how teachers interpret student thinking, value linguistic and cultural resources, and respond

instructionally to emerging understanding. By centering formative assessment within culturally and linguistically diverse STEM classrooms, the present review seeks to connect the aforementioned bodies of literature and clarify how equity-oriented assessment practices can support MLs' participation and learning. Moreover, in addition to mapping empirical trends, this review brings together formative assessment theory, sociocultural perspectives on learning, and equity-oriented pedagogies to examine how assessment practices shape opportunities to learn in STEM classrooms serving MLs. By identifying the trends and challenges in current practices and examining inclusive assessment strategies from literature, we hope to raise awareness of equity-oriented approaches to assessment for better support of the diverse needs of MLs.

RESEARCH METHOD

A systematic review is a method of synthesizing findings of studies that are chosen with a well-designed systematic search to answer a research question or deepen understanding of a specific issue (Page et al., 2021; Sterne et al., 2019). We followed the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) as guidance for this systematic review (Page et al., 2021). In the article "Five Steps to Conducting a Systematic Review," systematic review steps in Health Sciences were explained clearly (Khan et al., 2003). Those steps have been adopted for systematic reviews in educational sciences several times (e.g., John & Yunus, 2021; Silva-Laya et al., 2020; Zakaria et al., 2019). In this study, we have also followed those five steps (Khan et al., 2003) as explained in the following subsections. ChatGPT (GPT-5.3; OpenAI, 2026), an artificial intelligence–based language assistance tool, was used during manuscript preparation to support minor editorial revisions and improve clarity. All methodological decisions, analyses, and interpretations were conducted by the authors, who take full responsibility for the content of the manuscript.

Step 1: Framing Questions for a Review

The first step of the systematic review is clearly specifying the research question(s). As explained in previous sections, there was a need for a systematic review to understand how CLR formative assessment practices are implemented in K-12 STEM classrooms in the U.S., particularly in multilingual and multicultural settings. Therefore, we examined the trends and challenges in CLR formative assessment practices for MLs in K-12 STEM education, with a focus on equity and inclusivity.

Step 2: Identifying Relevant Work

For the second step, we chose six databases that index the most prominent journals involving this work, such as International Multilingual Research Journal, Journal of Science Teacher Education, Science and Children: APA PsycInfo (EBSCO), Education Database, Education Research Complete, Social Sciences Citation Index, Teacher Reference Center, and Education Resources Information Center (ERIC). After determining the databases, we searched for publications which are relevant to the research questions. The variation of keywords and Boolean operators we used for the search in different databases can be found in Table 1.

Table 1. Databases and Keywords Used for the Initial Search

Keywords	Databases	Hits
("Culturally responsive formative assessment practices" OR "linguistically responsive formative assessment practices" OR "culturally responsive pedagogy" OR "linguistically responsive pedagogy" OR "Culturally Responsive Science" OR "linguistically Responsive Science" OR "Culturally Responsive Technology" OR "linguistically Responsive Technology" OR "Culturally Responsive Engineering" OR "linguistically Responsive Engineering" OR "Culturally Responsive Mathematics" OR "linguistically Responsive Mathematics") AND	Social Sciences Citation Index	16
	Education Database	146

Keywords	Databases	Hits
("equitable STEM assessment practices" OR "inclusive STEM assessment practices" OR "equitable science assessment practices" OR "equitable mathematics assessment practices" OR "equitable technology assessment practices" OR "equitable engineering assessment practices" OR "inclusive science assessment practices" OR "inclusive mathematics assessment practices" OR "inclusive technology assessment practices" OR "inclusive engineering assessment practices") AND ("Multilingual learners" OR "Emergent bilinguals" OR "English Language Learners" OR "English as an additional language" OR "English as a Second Language" OR "Limited English Proficient" AND "STEM" OR "STEM education" OR "science technology engineering mathematics" OR "integrated-STEM") AND ("formative assessment" OR "classroom assessment" OR "in-class assessment" OR "instructional assessment" OR "diagnostic assessment") AND ("K-12" OR "primary education" OR "secondary education" OR "elementary" OR "middle school" OR "high school")	APA PsycInfo + Education Research Complete + ERIC + Teacher Reference Center	20

Note. The variations of keywords and Boolean operators were used for the search in given databases.

Step 3: Assessing the Quality of Studies

In this step, we put the criteria for eligibility and exclusion. Our preliminary search suggested that the time frame between 2010 and 2024 would be the most suitable for this study given key STEM-focused advancements initiated around 2010 in the U.S. Notable examples include President Obama's Educate to Innovate Campaign (The White House - Office of the Press Secretary, 2009), aimed at improving STEM education, the America COMPETES Reauthorization Act of 2010 (2011), emphasizing innovation and STEM education, and the 2010 National Education Technology Plan (United States Office of Educational Technology, 2010), which advanced technology integration in education. Therefore, starting the timeframe from 2010 aligned well with the scope of our systematic review. We included peer-reviewed journal articles and practitioner-oriented publications because the nature of the research question required us to understand the practices in the field. In Table 2, the criteria for eligibility and exclusion were provided.

Table 2. Criteria for the Eligibility and Exclusion

Criterion	Eligibility	Exclusion
Timeline	2010 - 2024 (August)	Published before 2010
Region	The United States (Authors can be out of the U.S., but the target population must be in the U.S.)	Out of the United States
Language	English	Non-English
Literature Type	Peer-reviewed Journal Articles & practitioner-oriented publications	Books, book chapters, conference proceedings, dissertation, etc.
Methodology	Qualitative, Quantitative, Mixed, systematic reviews, meta-analyses, meta-syntheses, conceptual/non-empirical approach	Policy recommendations/reviews, professional development papers, library research, etc.

Step 4: Summarizing the Evidence

We searched six databases using the keywords (see Table 1), applying the criteria for timeline, region, language, and literature type listed in Table 2. The elimination process is represented in Figure 1. The initial search yielded a total of 182 articles. We then removed duplicate articles that appeared in the results because

the keywords were set to search "anywhere in the article." After this step, 79 articles remained. Then, we reviewed carefully the abstracts of these articles and eliminated the irrelevant studies. At this point, we had 35 articles. Finally, we read all of them and decided to remove six articles because their focus was not formative assessment, STEM, and MLs, leaving a total of 25 articles, as represented in Figure 2. All these articles were downloaded and saved in our shared OneDrive folder for further analysis.

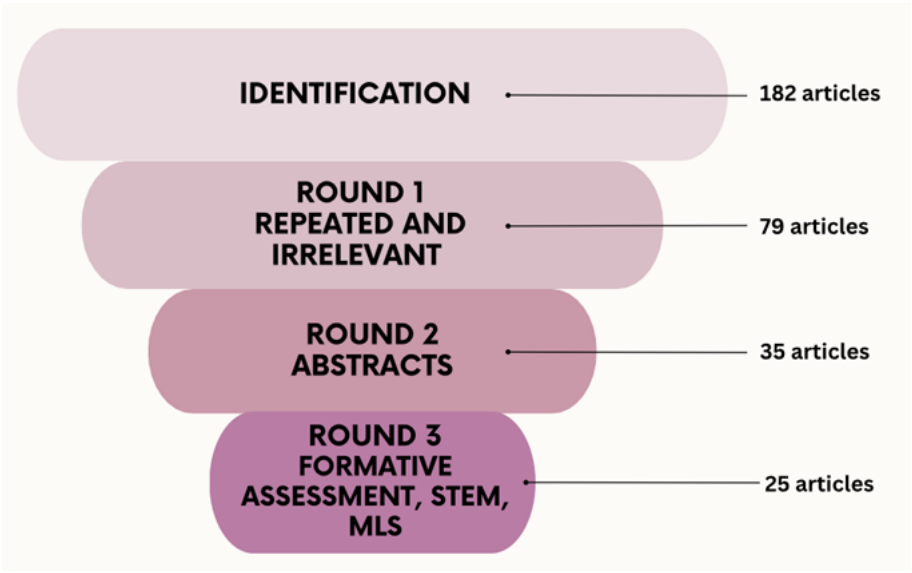

Figure 1. Elimination Process

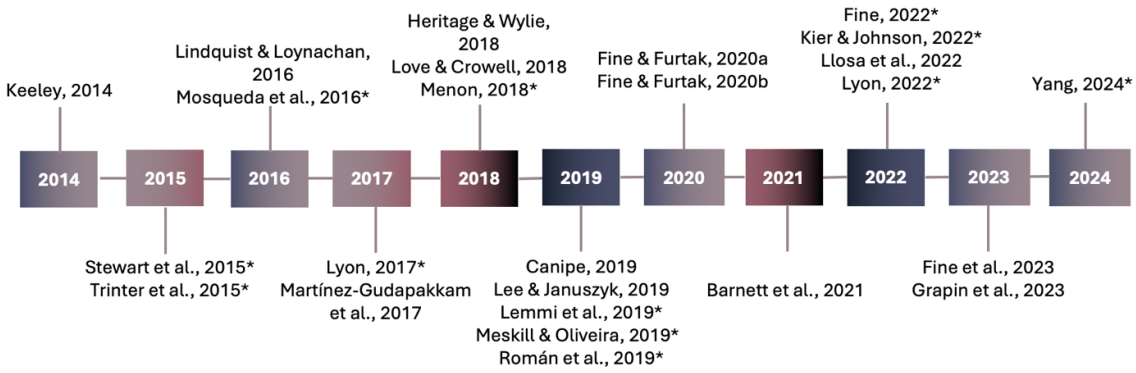

Figure 2. The Timeline of Reviewed Articles

Note. Articles with asterisks (*) are empirical studies.

Step 5: Interpreting the Findings

In this final step, 25 articles were reviewed: 13 non-empirical and 12 empirical, as shown in Figure 2 and Table 3. The empirical studies included 10 qualitative studies, 1 mixed-method study, and 1 multi-method study. Reviewed studies were published between 2014 and 2024. Further details about the articles can be seen in Table 3. Four authors reviewed all the articles and conducted thematic analysis (Braun & Clarke, 2006) guided by the research question: What are the key trends and challenges in CLR formative assessment practices for MLs in K-12 STEM education, with a focus on equity and inclusivity?

Table 3. Overview of the Articles

No	Author(s), Year	Methodology	Participants/Settings
1	Barnett et al., 2021	Conceptual/Non-empirical	First and fifth grade students in STEM learning at a bilingual program
2	Canipe, 2019	Conceptual/ Non-empirical	First grade students studying balance and motion in their science unit
3	Fine, 2022	Qual	A sixth-grade science teacher
4	Fine & Furtak, 2020a	Conceptual/Non-empirical	Framework for science classroom assessments for emergent bilingual learners
5	Fine & Furtak, 2020b	Conceptual/Non-empirical	Science assessment for emergent bilingual learners (SAEBL) Checklist
6	Fine et al., 2023	Conceptual/ Non-empirical	The (Trans)Formative Assessment Planning Template - how to incorporate Translanguaging in lesson planning
7	Grapin et al., 2023	Conceptual/ Non-empirical	A literature review
8	Heritage & Wylie, 2018	Conceptual/ Non-empirical	First and second grade students learning math
9	Keeley, 2014	Conceptual/ Non-empirical	Primary school context
10	Kier & Johnson, 2022	Qual	Four of the teacher–mentor pairs (three middle and one high school)
11	Lee & Januszyk, 2019	Conceptual/ Non-empirical	Fourth grade students learning science
12	Lemmi et al., 2019	Qual	19 middle or high school teachers
13	Lindquist & Loynachan, 2016	Conceptual/ Non-empirical	Fifth grade students learning science
14	Llosa et al., 2022	Conceptual/ Non-empirical	Fifth grade students learning science
15	Love & Crowell, 2018	Conceptual/ Non-empirical	Fourth and seventh grade special educators and language development specialists
16	Lyon, 2017	Qual + Quan	15 science teachers (three middle and 12 high school)
17	Lyon, 2022	Qual	Two middle school science teachers
18	Martínez-Gudapakkam et al., 2017	Conceptual/ Non-empirical	Fourth grade students learning science
19	Menon, 2018	Qual	Seventh grade science classroom and the teacher
20	Meskill & Oliveira, 2019	Qual	40 high school teachers
21	Mosqueda et al., 2016	Mixed	Eight mathematics teachers and 104 ninth graders
22	Román et al., 2019	Qual	Two science teachers (one elementary and one middle school)
23	Stewart et al., 2015	Qual	12 preservice teachers' submitted performance assessments (three in each of these subjects: science, mathematics, English language arts, history/social studies)
24	Trinter et al., 2015	Qual	Three second grade teachers and five second-grade students identified as mathematically promising
25	Yang, 2024	Qual	One fifth grade math teacher and seven emergent bilingual students

FINDING

In this section, we synthesize key trends and challenges in CLR formative assessment practices for MLs in K–12 STEM education, drawing on both empirical and non-empirical studies. In this review, trends refer to recurring, widely emphasized practices and approaches across studies, while challenges refer to persistent limitations, gaps, or barriers that constrain their implementation. Across both empirical and non-empirical literature, four interrelated dimensions of CLR formative assessment structure this synthesis:

(1) linguistic responsiveness,
(2) multimodal and inclusive assessment design,
(3) social and collaborative assessment processes, and
(4) teacher capacity and systemic conditions.

While these dimensions are analytically distinguished, they are closely interconnected in practice, with challenges shaping the extent to which identified trends can be enacted. The first part of this section presents these dimensions as recurring trends across studies, followed by a synthesis of cross-cutting challenges that influence their implementation.

Key Trends in CLR Formative Assessment: Emergent Dimensions

Linguistic Responsiveness in Assessment: Across the reviewed empirical studies, integrating students' linguistic resources into formative assessment emerged as a central and recurring approach in CLR STEM classrooms. Practices for language support in the reviewed studies included using simplified language in test items/communication, allowing for the use of home or everyday languages in answers and discussions, and incorporating student-generated language in rubrics (Fine, 2022; Lyon, 2017; Lyon, 2022; Román et al., 2019; Yang, 2024). Collectively, these practices reflect a shift toward positioning MLs' home languages as a resource for meaning-making rather than as a barrier to assessment participation.

A particularly prominent strand within this body of work is the use of translanguaging as an assessment practice. Across studies, translanguaging has been recognized as a valuable strategy for integrating students' linguistic repertoires into formative assessments in MMCs. For example, Fine (2022) demonstrated how translanguaging can be leveraged in science assessments, even when teachers do not share their students' linguistic repertoires. Through reflective collaboration, a teacher improved her translanguaging interpretive power, which led to a more linguistically responsive learning environment. Similarly, Lemmi et al. (2019) emphasized the role of teacher language ideologies, arguing that increased awareness of their own perspectives can foster more inclusive practices in MMCs. Yang (2024) further positioned translanguaging as a direct language support strategy, highlighting its effectiveness in responsive formative assessment accommodations that allow MLs to more accurately show their learning.

Translanguaging has also been framed as an asset-oriented approach to assessment. Lyon (2022) described assessments as "a tool to bring emergent bilinguals into a classroom learning community and scaffold their science learning and language/literacy development by continuously eliciting, recognizing, valuing, and then adapting to the linguistic and cultural assets that emergent bilinguals bring" (p. 229). In this view, translanguaging should be visible in assessment activities in MMCs as a means of embracing the assets MLs bring into the classroom. Together, these studies demonstrate that integrating translanguaging into formative assessments in MMCs can enhance student engagement, promote equitable learning opportunities for all students, and encourage teachers to adopt more inclusive assessment ideologies.

Besides translanguaging, several studies highlighted efforts to integrate language and content goals within assessment design. For example, Mosqueda et al. (2016) utilized an Adolescent Mathematics Writing Assessment System model which features an iterative cycle of student assessment, scoring and analyzing data, PD, coaching sessions, and informing instruction. Students' written responses to math questions were evaluated

using a rubric designed to align with grade-level standards for both mathematics and English language development. As part of the model, teachers were trained to provide daily math writing and math language routines. In the routine, students worked in small groups to correct grammatical errors embedded in math problems before solving them. Additionally, students read about Latina/o mathematicians to make learning more relevant and motivating for Latinx MLs. Similarly, Yang (2024) illustrated how structured classroom discussion supported both conceptual understanding and language development. These approaches underscore the role of assessment in STEM classrooms as a site where language and content learning are co-constructed rather than treated as separate domains.

Non-empirical literature further reinforces this emphasis on cultural and linguistic responsiveness by advocating for assessment practices that actively incorporate students' cultural and linguistic backgrounds. For example, Fine et al. (2023) shared a tool for science teachers of MLs to use to modify or create formative assessment tasks that build from their prior knowledge and welcome translanguaging, while Grapin et al. (2023) cautioned against teachers' restrictive views and use of language translations as a mere accommodation for MLs during assessment practices. Instead, teachers should "design assessment tasks that validate and sustain [MLs'] multiple linguistic and cultural ways of knowing" (Fine & Furtak, 2020a, p. 394). For instance, the teachers can create a learning space that welcomes their MLs' ideas in Spanish or English (Bartlett et al., 2021; Fine et al., 2023; Grapin et al., 2023; Keeley, 2014). In their example of using science notebooks with a group of fifth graders as a language and content learning and formative assessment tool, Lindquist and Loynachan (2016) saw that these notebooks made visible the students' linguistic and scientific knowledge and development. Moreover, Keeley (2014) emphasized the importance of culturally responsive formative assessment and shared how "small but important change[s]" (p. 33), such as names in task or assessment questions to reflect the cultures represented in MLs in a classroom, can "help students see that people of different cultures engage in science" (p. 33). Encouraging MLs to share their personal experiences and recognizing and utilizing their cultures and linguistic repertoires can lead to empowerment of MLs and asset-oriented perspectives toward them by the teachers and peers (Bartlett et al., 2021; Fine & Furtak, 2020b; Heritage & Wylie, 2018). Together, these studies point to a broader trend toward asset-oriented assessment practices that recognize and sustain the diverse ways in which MLs make meaning in STEM contexts.

Multimodal and Inclusive Assessment Design: Multimodal support in the assessment practices identified in the reviewed studies mainly focused on using multiple modes of assessment or multiple task types, such as verbal expression, written language, kinesthetic activities, stickers, numbers, and illustrations, which allow multiple ways for demonstrating understanding (Fine, 2022; Lyon, 2017; Menon, 2018; Román et al., 2019; Stewart et al., 2015; Trinter et al., 2015). For example, Trinter et al. (2015) employed a problem-based assessment incorporating multimodal approaches such as student drawings, written responses, and stickers. Their findings revealed that when presented with appropriately challenging questions, traditionally underrepresented students in mathematics were more engaged and demonstrated greater mathematical potential. Similarly, Fine (2022) described the co-design of a (trans)formative assessment that encouraged MLs to express their understanding of science concepts in both English and Spanish through written and oral tasks. Additionally, Lyon (2017) provided workshops to teachers on responsive formative assessment for MLs and then evaluated various aspects of the teachers' assessment practices, including their approaches to addressing the language demands of MLs. Common strategies included modifying text by incorporating visuals, glossaries, and bolded terms to enhance accessibility; assessing the same objective using multiple modes; implementing differentiated activities while maintaining consistent expectations; and promoting heterogeneous grouping for peer collaboration. These approaches suggest that multimodal design not only broadens access to assessment but also creates opportunities for deeper engagement with disciplinary practices.

Across studies, other strategies identified in teachers' practices include revoicing, restating, or paraphrasing students' responses to elicit their thinking, allowing extra wait time for MLs to respond, and

simultaneously transcribing key ideas of verbal discussions on the board. The availability of both spoken and written language provided valuable multimodal support for MLs. Meskill and Oliveira (2019) and Kier and Johnson (2022) further highlighted the use of multimodal tools in linguistically responsive formative assessment in MMCs. From digital technologies to analog resources, these tools can provide opportunities to scaffold learning for linguistically diverse students while addressing complex STEM concepts. As Fine (2022) and Yang (2024) framed the multimodal approaches as a form of language support in linguistically responsive classrooms, Lyon (2022) addressed how a partnership between a science teacher and a language teacher leveraged multimodal assessment practices to evaluate both language acquisition and content knowledge of students. Collectively, these studies suggested that leveraging multiple modes enhanced student engagement, accommodated diverse learning needs, and provided more equitable opportunities for MLs to demonstrate their learning.

Multimodal ways of learning and assessment strategies were essentially employed in all non-empirical articles we reviewed. In their work with fifth grade MLs in science instruction, Llosa et al. (2022) highlighted that providing opportunities to demonstrate MLs' learning in multiple ways is beneficial because the English language is not the only mode to express their understanding. They can show their thinking through drawings, orally, in English and in their home languages, individually and with their peers. Keeley (2014) explained that MLs may not recognize the vocabulary in English on a list, so providing pictures along with students' home language under the English word can provide insight into their thinking. Heritage and Wylie (2018) argued that MLs' mathematical explanations and justifications coming from developing English skills can be perceived as a lack of understanding of content. Instead, teachers should intentionally obtain evidence of learning through other modes of formative assessment, such as from dialogue, demonstration, and observation. Furthermore, Grapin et al. (2013) asserted that teachers ought not to privilege learning expressed exclusively through linguistic modalities, but through multimodal assessments, such as visual, written, and oral. Barnett et al. (2021) recommended providing opportunities for MLs to "demonstrate their knowledge in less language-embedded tasks and to express themselves through linguistic and nonlinguistic modes" (p. 58). Such experiences prevent MLs' English-language proficiency from masking their true understanding of the STEM content. To support science assessment designers "to represent and legitimize multiple ways of knowing and doing science" (Fine & Furtak, 2020b, p. 39), the SAEBL checklist can be employed. Fine & Furtak (2020b) developed this tool from a review of theoretical, empirical, and practitioner-focused work and explained in depth using a template how this tool can be used. Other multimodal formative assessment activities highlighted to check for understanding that will drive instruction are quick quizzes, exit tickets, entries in science and math journals, color-coding, sticky notes, diagrams/drawings, and taking photos (Love & Crowell, 2018; Martínez-Gudapakkam et al., 2017). Taken together, these findings position multimodal assessment as a central mechanism for decoupling language proficiency from the demonstration of disciplinary understanding, thereby expanding what counts as valid evidence of learning in STEM contexts.

Social and Collaborative Assessment Processes: Collaboration emerged as a key trend across almost all empirical studies, though its nature varied. Some studies emphasized teacher-researcher partnerships, while others focused on peer collaboration (among students and among teachers), assessment co-design, and mentorship structures to support MLs in MMCs. Fine (2022) highlighted the power of teacher-researcher collaboration, demonstrating how reflective conversations within the (Trans)Formative Assessment Co-design (TAC) model deepened a teacher's understanding of linguistically responsive practices. Notably, this collaborative reflection even led the teacher to develop new strategies for engaging families through translanguaging. Similarly, Mosqueda et al. (2016) found that coaching sessions for mathematics teachers in MMCs provided targeted support, reinforcing equitable assessment strategies.

Beyond teacher-researcher collaboration, peer mentoring and professional networks played a vital role in fostering inclusive formative assessment. Lyon (2022) emphasized that science teachers benefited

significantly from peer mentoring and shared reflection, enhancing their ability to implement inclusive assessment practices. Likewise, Meskill and Oliveira (2019) examined the co-design processes of language and content teachers, demonstrating that structured collaboration helped align language acquisition with STEM learning, ultimately improving both assessment practices and instructional strategies. Kier and Johnson (2022) further showed that teacher-undergraduate mentor partnerships in STEM settings, particularly when enhanced by digital tools, supported culturally responsive STEM learning and feedback mechanisms.

In addition to the strategies outlined above, peer collaboration was another strategy that was incorporated into different types of formative assessments, particularly informal formative assessments. A common example was group discussions. Román et al. (2019) examined the explanation practices of two science teachers as they engaged with students during informal assessments. They found that both teachers engaged in more frequent explanations with MLs compared to other students. However, their interaction patterns with MLs differed significantly. One teacher constantly engaged MLs in both whole-class and small-group discussions and created opportunities for them to work with mainstream students, while the other teacher only interacted with MLs in language-specific groups after whole-class discussions. Moreover, the first teacher continuously assessed MLs' understanding of target concepts throughout the lesson, whereas the second teacher confined assessments to the small-group sessions. These findings underscore the importance of consistently including MLs in the community of learners, fostering opportunities for social interaction and collaboration, a practice strongly aligned with sociocultural perspectives on learning. Peers serve as valuable resources in facilitating MLs' understanding of scientific concepts while simultaneously reducing anxiety through collaborative dialogue and shared problem-solving.

Non-empirical literature further emphasizes the role of discourse in formative assessment. For example, in Barnett et al.'s (2021) project-based learning unit designed for engaging young MLs (first and fifth graders) in coding, the teacher employed simple questioning throughout to informally assess her learners. Similarly, Canipe (2019) highlighted the use of simple questioning that created space for rich discussions in a 1st grade classroom in their study in which students experimented with making sense of scientific experiences by building explanations from evidence collected during projects. Discourse moves such as revoicing student ideas and connecting to their peers that fostered student-centered conversations were made by the teacher. Since STEM subjects are language-intensive for all students, Lee and Januszyk (2019) especially recommended teachers of MLs to give these students, who are developing both language and disciplinary knowledge, opportunities to engage in conversations about the topic at hand. Writing on the board, students' observations from discussions can also help the teacher assess their thinking (Canipe, 2019). Overall, these approaches create opportunities for students to articulate their ideas, engage with peers, and demonstrate understanding in ways that extend beyond traditional assessment formats.

Teacher Capacity, Feedback, and Systemic Conditions: Across the reviewed studies, teacher capacity emerged as a critical factor in the implementation of CLR formative assessment practices. Several studies, including Fine (2022), Lemmi et al. (2019), Lyon (2022), Meskill and Oliveira (2019), and Stewart et al. (2015) repeatedly underscored the critical role of teachers' self-reflection in adapting CLR instructional and assessment strategies. While some studies emphasized its role in addressing implicit biases and power dynamics in the classroom regarding race and language and promoting culturally responsive teaching (e.g., Lemmi et al., 2019; Lyon, 2022), others positioned teacher reflection as a necessary component of instruction and assessment design in MMCs (e.g., Fine, 2022; Stewart et al., 2015). Meskill and Oliveira (2019) further expanded this idea by highlighting the value of collaborative reflection between content and language teachers, emphasizing the importance of interdisciplinary dialogue in fostering more inclusive assessment practices.

In addition to reflection, targeted assessment practices such as providing tailored feedback were emphasized as important components of culturally responsive teaching. Kier and Johnson (2022) explored its role in teacher-undergraduate mentor collaborations during COVID-19 school closures, revealing that providing

individualized feedback and allowing students time to reflect fostered a culturally responsive learning environment. Similarly, Stewart et al. (2015) highlighted strategies used by preservice teachers to provide feedback on their students' understanding, such as drawing science concept illustrations and filling out Know-Want to Know-Learned charts. Beyond teacher-driven feedback, Lyon (2022) touched upon the value of peer and self-feedback in helping students deepen their understanding of scientific concepts while embracing their cultural and linguistic backgrounds. Yang (2024) noted that students expressed their desire for tailored feedback, as well as enhanced language support like access to a vocabulary glossary, highlighting areas for improvement. Overall, these studies suggest that tailored feedback can enhance both STEM content comprehension and culturally responsive teaching practices in MMCs. These practices also highlight the role of assessment as an ongoing, interactive process that responds to students' diverse learning needs.

Cross-Cutting Challenges in CLR Formative Assessment

Uneven Implementation and Persistent Monolingual Norms: Despite the strong emphasis on linguistic responsiveness, the reviewed studies also revealed substantial challenges in its consistent implementation. Significant disparities were observed in the degree of implementation across studies. For example, while various translanguaging strategies have been utilized in empirical research (e.g., Fine, 2022; Lyon, 2022; Yang, 2024), they were not often explicitly or systematically designed to leverage MLs' full linguistic repertoire, funds of knowledge, or prior experiences (García & Lin, 2017; Vélez-Ibáñez, 1988).

Moreover, across empirical studies, a persistent structural constraint is the dominance of monolingual norms in STEM education. Lemmi et al. (2019) and Mosqueda et al. (2016) revealed how rigid language standards and embedded monolingual ideologies in educational settings can marginalize MLs, hindering their ability to demonstrate content knowledge and engage meaningfully in assessments. In Lyon (2017), only 34% of classroom observations applied language support for MLs during assessments. Furthermore, no teachers were found to integrate students' funds of knowledge or design assessments that connected meaningfully to the lived experiences of MLs. Yang (2024) found limited evidence that the teacher explicitly leveraged students' language backgrounds, such as their English proficiency and home languages, in the assessment practices, which aligns with the findings of Lyon (2022) about "missed opportunities for students to use a repertoire of linguistic resources" (p. 24). Lyon (2022) argued that addressing this challenge requires systemic shifts in how both language proficiency and content learning are conceptualized and assessed in science classrooms.

In addition, balancing language development with disciplinary rigor remains an ongoing challenge. Lemmi et al. (2019), Meskill and Oliveira (2019), and Román et al. (2019) highlighted the challenge of balancing STEM content instruction with language acquisition support. They noted that teachers often struggle to design assessments that effectively address both language development and content mastery, particularly when working with students of varying linguistic and academic needs and proficiencies. Similarly, Lyon (2022) pointed to the difficulty of integrating disciplinary literacy into assessments without ignoring the value of MLs' linguistic backgrounds. This tension reflects a broader disconnect between the theoretical emphasis on integrating language and content and the practical realities of classroom assessment.

Taken together, these findings suggest that while linguistic responsiveness is widely recognized as a critical component of equitable formative assessment, its enactment is shaped by both teacher-level constraints and broader systemic conditions that continue to privilege standardized, monolingual forms of assessment.

Gaps in PD and Teacher Preparation: Despite the recognized importance of teacher capacity, the reviewed studies highlighted the inadequacy of PD in addressing the linguistic demands of STEM instruction and assessment, particularly in integrating language development with disciplinary content. For instance, Lemmi et al. (2019) argued that current PD programs often fail to integrate the linguistic aspects into science education and underlined the need to discuss language ideologies in science instruction to evaluate how

different approaches impact the learning of students with varying language and academic proficiencies. Meskill and Oliveira (2019) also addressed the lack of PD as a root cause of challenges experienced by MLs and their teachers in science classrooms, underlining that many educators are underprepared to design instruction and assessments that align with both language needs and content goals. Stewart et al. (2015) and Lyon (2017) further emphasized the lack of sustained and targeted PD as one of the major barriers. Taken together, these findings indicate that advancing CLR formative assessment is not solely a matter of adopting new strategies but also requires sustained investment in teacher learning and systemic supports that enable teachers to enact CLR practices in coherent and integrated ways.

DISCUSSION

This review offers several implications for understanding and advancing CLR formative assessment in K–12 STEM education. Rather than identifying a fixed set of effective practices, the findings suggest that formative assessment for MLs is shaped by how multiple dimensions of assessment are coordinated in practice. First, the findings indicate that CLR formative assessment should be understood as a systemic process rather than a collection of individual strategies. Prior research describes formative assessment as involving the elicitation, interpretation, and instructional use of student thinking (Heritage & Wylie, 2018; Llosa et al., 2022). This review builds on that understanding by examining how these processes are enacted in culturally and linguistically diverse STEM classrooms, where they intersect with linguistic, pedagogical, and social dimensions. When these elements are not aligned, formative assessment practices may remain fragmented and not fully support equitable learning.

Second, the findings highlight teacher interpretation as a critical leverage point in CLR formative assessment. While formative assessment is often framed in terms of eliciting student responses and providing feedback, its effectiveness depends on how teachers interpret student thinking. In STEM classrooms, this interpretation requires attention to multiple forms of communication, including talk, text, symbols, and visual representations (Lee & Januszyk, 2019). This suggests that supporting teachers in developing interpretive expertise is essential for recognizing and responding to MLs' understanding in meaningful ways.

Third, the findings point to the need to reconsider the role of language in STEM assessment. Assessment practices often address linguistic demands through simplification or basic scaffolding to make tasks more accessible. However, the reviewed studies suggest that these approaches may be limited and emphasize the importance of providing language-rich opportunities for students to demonstrate understanding through diverse linguistic and multimodal resources, including translanguaging (Fine et al., 2023; Grapin et al., 2023; Lyon, 2017). This suggests that language is not simply a barrier to be addressed, but a central component of how knowledge is constructed and assessed in STEM contexts.

Fourth, the findings highlight a persistent gap between the identification of promising practices and their consistent enactment in classrooms. While a range of approaches has been proposed to support MLs, their implementation depends on contextual factors such as time, resources, and access to professional learning. Similar concerns about variability in implementation and the conditions required for effective use have been noted in the formative assessment literature (Bennett, 2011).

While the reviewed studies consistently emphasize the value of CLR and multimodal assessment practices, there remains a notable gap between theoretical recommendations and classroom enactment. This tension suggests that the challenge is not only pedagogical but also structural, as teachers operate within systems that continue to prioritize standardized and monolingual forms of assessment. This may explain why, despite strong conceptual support, the implementation of CLR formative assessment remains uneven across contexts.

Taken together, these implications point to the need for a shift in how CLR formative assessment is conceptualized and supported. Efforts to improve assessment practices should move beyond the adoption of individual strategies and instead focus on developing integrated approaches that connect elicitation, interpretation, and instructional response. In addition, professional learning should support teachers in interpreting multilingual and multimodal forms of student thinking, while assessment design should recognize language as central to disciplinary learning rather than as a constraint.

Finally, this review is subject to several limitations. First, the use of the term MLs may obscure important differences in students' linguistic, cultural, and educational backgrounds. Second, the inclusion of both empirical and non-empirical studies provides a broader perspective on the topic, but may limit the generalizability of the findings. Third, the reviewed studies vary in context, grade level, and disciplinary focus, which makes it difficult to draw conclusions about how CLR formative assessment operates in specific settings. These limitations suggest the need for more context-specific and empirically grounded research that examines how these practices are enacted across different STEM disciplines and grade levels.

CONCLUSION

This systematic review suggests that CLR formative assessment in K–12 STEM education has been shaped by the interaction of multiple dimensions, including linguistic, multimodal, social, and teacher-related factors within broader systemic contexts. These findings indicate that advancing equitable assessment for MLs requires moving beyond isolated strategies toward more integrated and context-sensitive approaches. In particular, strengthening teacher capacity, recognizing multilingual and multimodal forms of meaning-making, and addressing structural constraints are critical for supporting meaningful and sustained change. By identifying key patterns and tensions in current practice, this review contributes to ongoing efforts to develop more equitable and CLR formative assessment practices in STEM education. Beyond its immediate scope, this review contributes to current international discussions on equity in STEM education by highlighting how assessment practices can either constrain or expand opportunities for MLs. These insights may inform future research, policy development, and instructional design across diverse educational systems.

ACKNOWLEDGMENTS

Bilgehan Ayik acknowledges that her graduate education has been sponsored by the Ministry of National Education in Türkiye. This study does not necessarily reflect the views of her sponsor.

REFERENCES

Aguirre-Muñoz, Z., Pando, M., & Liu, C. (2024). Enhancing bilingual/ESL teachers' STEM instruction with targeted content and disciplinary literacy professional development: A study on knowledge and practice outcomes. *Education Sciences, 14*(7), 745. https://doi.org/10.3390/educsci14070745

America COMPETES Reauthorization Act of 2010, Pub. L. No. 111–358, 124 Stat. 3982 (2011). https://www.congress.gov/111/plaws/publ358/PLAW-111publ358.pdf

Ayik, B., Kim, S., Zan, Y., Gu, D., & Kim, W. L. (2026). Supporting multilingual learners in K–12 STEM classrooms through formative assessments: A qualitative descriptive approach. *Natural Sciences Education, 55*, e70048. https://doi.org/10.1002/nse2.70048

Bailey, F., Burkett, B., & Freeman, D. (2008). The mediating role of language in teaching and learning: A classroom perspective. In B. Spolsky & F. M. Hult (Eds.), *Handbook of educational linguistics* (pp. 606–625). Blackwell. https://sabinemendesmoura.wordpress.com/wp-content/uploads/2014/11/the_handbook_of_educational_linguistics.pdf

Barnett, E., Perez, K. M., & Delgado, R. (2021). Que es un codigo? Science and Children, 58(5), 51-58. https://www.jstor.org/stable/pdf/27133305.pdf

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. DOI: https://doi.org/10.1080/0969595980050102

Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971) *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. DOI: https://doi.org/10.1191/1478088706qp063oa

Bruner, J. (1985). *Actual minds, possible worlds*. Harvard University Press.

Buxton, C. A., & Lee, O. (2014). English language learners in science education. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (pp. 204–222). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780203097267-20/general-instructional-methods-strategies-david-treagust-chi-yan-tsui

Canipe, M. (2019). What Did You Notice?: Simple questions lead to evidence and explanations. *Science and Children, 57*(4), 32–35. https://www.jstor.org/stable/26901561

Cheung, K. K. C., & Ng, D. T. K. (2026). A systematic review of research on translanguaging in STEM education. *International Journal of Multilingualism, 23*(1), 28–45. https://doi.org/10.1080/14790718.2025.2512783

Congressional Research Service. (2018). Science, Technology, Engineering, and Mathematics (STEM) Education: An Overview. https://crsreports.congress.gov/product/pdf/R/R45223

Corntassel, J., & Hardbarger, T. (2019). Educate to perpetuate: Land-based pedagogies and community resurgence. *International Review of Education, 65*, 87–116. DOI: https://doi.org/10.1007/s11159-018-9759-1

Fine, C. G. M. (2022). Translanguaging interpretive power in formative assessment co-design: A catalyst for science teacher agentive shifts. *Journal of Language, Identity & Education, 21*(3), 191–211. DOI: https://doi.org/10.1080/15348458.2022.2058858

Fine, C. G. M., & Furtak, E. M. (2020a). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education, 104*(3), 393–420. DOI: https://doi.org/10.1002/sce.21565

Fine, C. G. M., & Furtak, E. M. (2020b). The SAEBL Checklist. *The Science Teacher, 87*(9), 38–48. https://www.jstor.org/stable/pdf/27048185.pdf

Fine, C. G. M., Littich, H., & Getz, M. (2023). The (trans) formative assessment planning template. *Science Scope, 46*(31). https://www.nsta.org/science-scope/science-scope-januaryfebruary-2023/transformative-assessment-planning-template

Furtak, E., Pasquale, M., & Aazzerah, R. (2016, January). *How teachers can develop formative assessments that fit a three-dimensional view of science learning*. STEM Teaching Tools. https://stemteachingtools.org/assets/landscapes/STEM-Teaching-Tool-18-3D-Formative-Assessment.pdf

García, O., & Lin, A. M. (2017). Translanguaging in bilingual education. In O. García, A. M. Lin, & S. May (Eds.), *Bilingual and Multilingual Education* (pp. 117–130). DOI: https://doi.org/10.1007/978-3-319-02324-3_9-1

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.

Gay, G. (2021). Culturally responsive teaching: Ideas, actions, and effects. In H. R. Milner & K. Lomotey (Eds.), *Handbook of urban education* (pp. 212–233). Routledge.

Grapin, S. E., Pierson, A., González-Howard, M., Ryu, M., Fine, C., & Vogel, S. (2023). Science education with multilingual learners: Equity as access and equity as transformation. *Science Education, 107*(4), 999–1032. DOI: https://doi.org/10.1002/sce.21791

Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Paper prepared for the Council of Chief State School Officers. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://files.eric.ed.gov/fulltext/ED543063.pdf

Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: Achievement, identity, and equity. *ZDM – Mathematics Education, 50*(4), 729–741. DOI: https://doi.org/10.1007/s11858-018-0943-3

John, E., & Yunus, M. M. (2021). A systematic review of social media integration to teach speaking. *Sustainability, 13*(16), 9047. DOI: https://doi.org/10.3390/su13169047

Keeley, P. (2014). Formative assessment probes: Assessment for all. *Science and Children, 51*(5), 32–35. www.jstor.org/stable/43176531

Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine, 96*(3), 118–121. DOI: https://doi.org/10.1177/014107680309600304

Kier, M. W., & Johnson, L. L. (2022). Exploring how secondary STEM teachers and undergraduate mentors adapt digital technologies to promote culturally relevant education during COVID-19. *Education Sciences, 12*(1), 48. DOI: https://doi.org/10.3390/educsci12010048

Kim, S., Zhao, M., Kim, W., Ayik, B., Gu, D., Chen, X., Zan, Y., & Ramos, K. A. (2026). Supporting Multilingual Learners Through Translanguaging Pedagogy in U.S. K–12 STEM Classrooms: A Systematic Meta-Synthesis. *Education Sciences, 16*(3), 376. DOI: https://doi.org/10.3390/educsci16030376

Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into Practice, 34*(3), 159–165. https://doi.org/10.1080/00405849509543675

Lee, O., & Januszyk, R. (2019). Formative assessment of English language proficiency in the science classroom. *Science and Children, 56*(9), 80–85. https://www.jstor.org/stable/pdf/26901480.pdf

Lemmi, C., Brown, B. A., Wild, A., Zummo, L., & Sedlacek, Q. (2019). Language ideologies in science education. *Science Education, 103*(4), 854–874. DOI: https://doi.org/10.1002/sce.21508

Lindquist, B., & Loynachan, C. (2016). Learning science in a second language. *Science and Children, 54*(3), 47–51. DOI: https://doi.org/10.2505/4/sc16_054_03_47

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher, 35*(7), 3–12. https://doi.org/10.3102/0013189X035007003

Llosa, L., Grapin, S. E., & Haas, A. (2022). Assessment for English Learners. *Science and Children, 59*(3), 58–63. https://www.nsta.org/science-and-children/science-and-children-januaryfebruary-2022/assessment-english-learners

Louie, N. (2018). Culture and ideology in mathematics teacher noticing. *Educational Studies of Mathematics, 97*, 55–69. DOI: https://doi.org/10.1007/s10649-017-9775-2

Love, N., & Crowell, M. (2018). Strong teams, strong results. *The Learning Professional, 39*(5), 34–39. https://learningforward.org/wp-content/uploads/2018/10/strong-teams-strong-results.pdf

Lyon, E. G. (2017). Exploring secondary science teachers' enactment of assessment practices to reflect responsive science teaching for English learners. *Journal of Science Teacher Education, 28*(8), 674–698. DOI: https://doi.org/10.1080/1046560X.2017.1401415

Lyon, E. G. (2022). Reframing formative assessment for emergent bilinguals: Linguistically responsive assessing in science classrooms. *Science Education, 107*(1), 203–233. https://doi.org/10.1002/sce.21760

Martínez-Gudapakkam, A., Mutch-Jones, K., & Hicks, J. (2017). Methods and strategies: Formative assessment practices to support students who struggle in science. *Science and Children, 55*(2), 88–93. DOI: https://doi.org/10.2505/4/sc17_055_02_88

Menon, P. (2018). Role of assessment conversations in a technology-aided classroom with English language learners: An exploratory study. *Multicultural Education, 25*(2), 42–50. https://files.eric.ed.gov/fulltext/EJ1181627.pdf

Meskill, C., & Oliveira, A. W. (2019). Meeting the challenges of English learners by pairing science and language educators. *Research in Science Education, 49*, 1025–1040. DOI: https://doi.org/10.1007/s11165-019-9837-9

Mosqueda, E., Bravo, M., Solís, J., Maldonado, S. I., & De La Rosa, J. (2016). Preparing middle school students for the transition to high school mathematics: Assessing Latinas/os' mathematical understanding, academic language and English proficiency. *Bilingual Review/Revista Bilingüe, 33*(2). https://bilingualreviewjournal.org/index.php/br/article/view/258

National Academies of Sciences, Engineering, and Medicine. (2024). *Equity in K-12 STEM education: Framing decisions for the future*. The National Academies Press. DOI: https://doi.org/10.17226/26859

National Center for Education Statistics. (2024). *English learners in public schools: Condition of Education*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cgf/english-learners

National Research Council. (2012). *A Framework for K-12 science education: Practices, crosscutting concepts, and core Ideas.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*(2), 149–170. DOI: https://doi.org/10.1080/09695940701478321

Ntumi, S., Adzifome, S. N., Nyamekye, T., & Vedor, F. K. (2026). Culturally responsive assessment in mathematical word problems and numerical cognition in multilingual education. *Scientific Reports, 16*, 5133. https://doi.org/10.1038/s41598-026-35864-0

Olivera-Aguilar, M., Rikoon, S. H., & Hamilton, L. S. (2024). *Culturally responsive education and assessments in U.S. schools*. American Institutes for Research. https://www.air.org/sites/default/files/2024-04/culturally-responsive-education-assessment-US-schools.pdf

OpenAI. (2026). *ChatGPT (GPT-5.3)* [Large language model]. https://chat.openai.com/

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery, 88*, 105906. DOI: https://doi.org/10.1016/j.ijsu.2021.105906

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher 41*(3), 93–97. DOI: https://doi.org/10.3102/0013189X12441244

Piaget, J. (1954). *The construction of reality in the child*. Basic Books.

Ramírez, P. C. (2023). Reframing dual language education in the U.S*. International Journal of Bilingual Education and Bilingualism, 26*(7), 769–773. DOI: https://doi.org/10.1080/13670050.2022.2087469

Ramírez, P., & C. Faltis. (2020). *Dual language education in the US: Rethinking pedagogy, curricula, and teacher education to support dual language learning for all*. Routledge.

Robinson-Jones, C., Duarte, J., Chik, A., & Wang, H. C. (2026). Linguistically responsive assessment for learning in primary education: A scoping review. *Journal of Multilingual and Multicultural Development*, 1–29. https://doi.org/10.1080/01434632.2026.2613680

Román, D., del Rosal, K., & Basaraba, D. (2019). Constructing explanations in science: Informal formative assessment practices among science teachers of English Learners. *Research in Science Education, 49*(4), 1055–1067. DOI: https://doi.org/10.1007/s11165-019-9849-5

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gange, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. I, pp. 39–83). Rand McNally.

Silva-Laya, M., D'Angelo, N., García, E., Zúñiga, L., & Fernández, T. (2020). Urban poverty and education. A systematic literature review. *Educational Research Review, 29*, 100280. DOI: https://doi.org/10.1016/j.edurev.2019.05.002

Sterne, J. A., Hernán, M. A., McAleenan, A., Reeves, B. C., & Higgins, J. P. (2019). Assessing risk of bias in a non-randomized study. In J. P. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (pp. 621–641). Wiley. DOI: https://doi.org/10.1002/9781119536604.ch25

Stewart, A. R., Scalzo, J. N., Merino, N., & Nilsen, K. (2015). Beyond the criteria: Evidence of teacher learning in a performance assessment. *Teacher Education Quarterly, 42*(3), 33–58. https://www.jstor.org/stable/pdf/teaceducquar.42.3.33.pdf

Suh, E. K., Hoffman, L., & Zollman, A. (2020). STEM inclusion research for English Language Learners (ELLs): Making STEM accessible to all. In C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & L.

D. English (Eds.), *Handbook of research on STEM education* (pp. 311–322). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780429021381-29/stem-inclusion-research-english-language-learners-ells-emily-suh-lisa-hoffman-alan-zollman

Stephens, A., & Francis, D. (2018). *English learners in STEM subjects: Transforming classrooms, schools, and lives.* National Academies Press.

The White House - Office of the Press Secretary. (2009, November 23). *President Obama launches "Educate to Innovate" campaign for excellence in science, technology, engineering & math (STEM) education*. https://obamawhitehouse.archives.gov/the-press-office/president-obama-launches-educate-innovate-campaign-excellence-science-technology-en

Trinter, C. P., Moon, T. R., & Brighton, C. M. (2015). Characteristics of students' mathematical promise when engaging with problem-based learning units in primary classrooms. *Journal of Advanced Academics, 26*(1), 24–58. DOI: https://doi.org/10.1177/1932202X14562394

Trumbull, E., & Lash, A. (2013). *Understanding formative assessment. Insights form Learn theory Measurement theory.* WestEd. https://www2.wested.org/www-static/online_pubs/resource1307.pdf

United States Office of Educational Technology. (2010) *Transforming American education: learning powered by technology: National Education Technology Plan*. Washington, D.C.: U.S. Dept. of Education, Office of Educational Technology. https://www.loc.gov/item/2023692186/

Vélez-lbáñez, C. G. (1988). Networks of exchange among Mexicans in the U.S. and Mexico: Local level mediating responses to national and international transformations. *Urban Anthropology, 17*(1), 27–51. https://www.jstor.org/stable/40553124

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Harvard University Press.

White House Office of Science and Technology Policy. (2023). *2022 progress report on the implementation of the federal science, technology, engineering, and mathematics (STEM) education strategic plan*. https://files.eric.ed.gov/fulltext/ED630703.pdf

Williams, A. P., Athanases, S. Z., Higgs, J., & Martinez, D. C. (2020). Developing an inner witness to notice for equity in the fleeting moments of talk for content learning. *Equity & Excellence in Education, 53*(4), 504–517. DOI: https://doi.org/10.1080/10665684.2020.1791282

WIDA. (2020). *WIDA English language development standards framework, 2020 edition: Kindergarten–grade 12*. Board of Regents of the University of Wisconsin System. https://wida.wisc.edu/sites/default/files/resource/WIDA-ELD-Standards-Framework-2020.pdf

Wolf, M. K., Sova, L., Janssen, G., López, A. A., Gooch, R. M., Pooler, E., & Lee, J. (2025). Equity for multilingual learners: Leveraging formative assessment and socioculturally responsive assessment principles. *Bilingual Research Journal*, 1–21. https://doi.org/10.1080/15235882.2025.2492675

Yang, X. (2024). Linguistically responsive formative assessment for emergent bilinguals: Exploration of an elementary teacher's practice in a math classroom. *International Multilingual Research Journal, 19*(1), 67–90. DOI: https://doi.org/10.1080/19313152.2024.2339757

Zakaria, M. I., Maat, S. M., & Khalid, F. (2019). A systematic review of problem-based learning in education. *Creative Education, 10*(12), 2671–2688. DOI: https://doi.org/10.4236/ce.2019.1012194