Rethinking WIDA's ACCESS Test Validity and Reliability: Bibliometric and Content Analysis (2005-2025)

ISSN (Online): 3007-1038
Pages: 91–113
DOI: 10.55737/rl 2025.44132

https://regionallens.com

Housseine Bachiri [1]

Abstract: This study endeavors to investigate the validity and reliability of WIDA's ACCESS test since 2005 to present time, with a particular emphasis on its potency to accurately measure English language proficiency for English learners (ELs). Using a combination of bibliometric and content analysis methods, the research synthesizes insights from more than 30 scholarly and empirical studies, highlighting trends, methodological approaches, and key findings pertinent to ACCESS. In addition, the study analyzes ACCESS test data gathered across 14 Illinois school districts, during the 2022–2024 academic years in order to provide a current and data-driven perspective on student performance patterns. Moreover, the analysis centers on the Rasch Model, which underpins the ACCESS scoring system, providing a framework for measuring proficiency while also considering its inherent limitations. The findings disclose a complex interplay between systemic factors, such as the bilingual program structure (90:10 versus 50:50), instructional core, including teaching strategies and classroom supports, all of which collectively shape ACCESS test outcomes and interpretations. Similarly, the findings indicate that more longitudinal empirical research should be done so as to examine and evaluate the validity and reliability of ACCESS test, while simultaneously proposing a new practical testing philosophy that would theoretically and empirically benefit from the previously existing English language proficiency tests, and hence lead to more valid and reliable testing methods.

Key Words: ACCESS Test, Validity; Reliability, Psychometrics, Rasch Model, Bibliometric Analysis, Content Analysis, English Language Proficiency

Introduction

The impetus for this paper purely emanates from the researcher's profound interest in Applied Linguistics and Second Language Acquisition, particularly in ferreting out how WIDA's ACCESS test shapes the experiences of English Learners (ELs). In my daily work, I routinely analyze ACCESS data, emphasizing students' overall scale and composite scores, specifically (6-12) in order to measure English Language proficiency for placement (ESL classes) and instructional purposes, yet I am often disheartened by how many capable high school students fail to earn the Seal of Biliteracy because they fall short of the 4.8 score required to exit the EL program. This observation instantly prompted me to investigate the underlying factors influencing these outcomes, including the Teaching for Biliteracy Framework (Beeman & Urow, 2013), the quality of instruction, curricula, bilingual program structure (90:10 versus 50:50), and the ACCESS test itself. My research journey began with a close examination of the Teaching for Biliteracy Framework (Bachiri, 2025)—an approach adopted by several Illinois districts, including District 60—and continued with an extensive review of the evolution of the ACCESS test since its inception in 2005.

Through this process, the researcher conducted a comprehensive literature review focusing on the validity and reliability of WIDA's ACCESS test. This review explored how test design, scoring procedures, and alignment with instructional practices potentially contribute to the persistent challenges ELs face in meeting proficiency standards. The researcher also examined studies on the effectiveness of instructional frameworks, such as the Teaching for Biliteracy model, and how factors like differentiated instruction, teacher training, culturally responsive curricula, and classroom practices can support—or hinder—student performance on high-stakes assessments, raising a central question frequently posed by educators and administrators: Why do EL students fail the ACCESS test despite taking it each year multiple times?—or even more strikingly, why do some students fail the test even though they have been in the U.S. since kindergarten? These two questions are the backbone of this research paper.

[1] Bilingual & Multicultural Department, Lincoln Center-Illinois, Waukegan, Illinois, United States.
Email: housseinebachiri87@gmail.com

Citation: Bachiri, H. (2025). Rethinking WIDA's ACCESS Test Validity and Reliability: Bibliometric and Content Analysis (2005-2025). *Regional Lens, 4*(4), 91-113. https://doi.org/10.55737/rl.2025.44132

Corresponding Author: Housseine Bachiri (✉ housseinebachiri87@gmail.com)

There is no doubt that assessing English language proficiency among English learners (ELs) in U.S. schools has become a contentious and punctilious process with significant consequences for academic placement, curriculum design, and program exit. Researchers and critics have constantly shown that standardized tests frequently produce misleading inferences about students, teachers, and schools. According to Belinder et al. "we documented the distortion of instructional values when teachers focused on "bubble" kids—those on the cusp of passing the test—at the expense of the education of very low or very high scorers (Belinder et al., 2007, p. 3)." When teachers concentrate mainly on students who are just barely likely to pass a test ("bubble" kids), it can distort their teaching priorities. As a result, students who are struggling a lot or those who are excelling may get less attention. In the same vein, Koretz cautions that "tests themselves are not valid or invalid. Rather, inferences based on test scores are valid or not." (Koretz, 2008, p. 1). Nichols and Berliner argue that "the overvaluation of this single indicator of school success often compromises the validity of the test scores themselves." (Nichols & Berliner, 2007, p. 1). Linn clarifies the policymaking limits of score-based accountability, noting that reliance on single scores creates "severe limitations" for trustworthy inferences and can distort instruction and curriculum (2000, pp. 4–16).

The ACCESS test, developed by the WIDA Consortium in 2005 as part of the No Child Left Behind Act of 2001 (Illinois State Board of Education, 2005) has become the primary standardized assessment tool for measuring English proficiency across listening, speaking, reading, and writing domains for multilingual learners (MLs) (Patterson & Schneider, 2024, p. 3). Nevertheless, a pilot study conducted by the same researchers highlights that many ESOL teachers question the alignment between ACCESS scores and students' actual classroom performance. "Teachers reported that students with lower ACCESS scores were nonetheless able to engage meaningfully in oral discussions, complete complex assignments, and comprehend grade-level content, suggesting a disconnect between the standardized assessment and authentic language use (Patterson & Schneider, 2024, p. 5)." Additionally, Belinder et al. state that "dozens of assessment experts have argued eloquently and vehemently that the high-stakes tests accompanying the implementation of the No Child Left Behind Act are psychometrically inadequate for the decisions that must be made about students, teachers, and schools (2007, p. 3)." These findings echo broader critiques in the literature indicating that standardized ELP assessments often privilege academic forms of English and may fail to capture the context-dependent, nuanced language skills that MLs demonstrate in real-life communicative situations (Shohamy, 2001, p. 72).

More importantly, educators have expressed substantial concerns, as numerous surveys and interviews reveal that a major issue lies in the test's poor alignment with students' demonstrated abilities and achievements in the classroom (approximately 75% of teachers), student anxiety (60%; Solano-Flores & Trumbull, 2003), limited evidence of validity (55%; Echevarria, Vogt, & Short, 2017), digital administration challenges (50%; WIDA, 2020), and potential cultural bias in test items (45%; Abedi & Lord, 2001). This data suggests that while ACCESS provides standardized and comparative information, there are systematic concerns about its alignment with students' actual language abilities and learning experiences.

Test anxiety emerges as a significant factor affecting language performance. Aydin identified multiple sources of anxiety among foreign language learners, including physical, test-related, and affective problems, and suggested that teachers raise awareness of these issues and implement strategies to reduce anxiety (Aydin, 2020, p. 6). In the same manner, Zheng and Cheng (2018, p. 13) further demonstrated that cognitive test anxiety was a significant negative predictor of language achievement, underlining the impact of anxiety on standardized test performance. In digital assessment contexts, Deng emphasized the interplay between techno-competence, teacher support, self-efficacy, autonomy, and test anxiety, highlighting the importance of addressing these factors to reduce anxiety in online assessments (2025, p. 12).

Furthermore, Foster argued that such testing disproportionately affects marginalized communities by narrowing the curriculum, limiting teacher autonomy, and prioritizing test scores over holistic educational approaches (Foster, 2024, p. 5). This perspective aligns with research emphasizing the role of language anxiety in dictating learning outcomes. Similarly, Liu found that foreign language anxiety could act as a barrier to effective learning for ELs, even when curiosity positively influenced language acquisition (2025, p. 4).

Research Questions

1. To what extent do ACCESS scores reflect English learners' ability to use English in meaningful and real-world classroom tasks?
2. How does test design (e.g., multiple-choice items, prompts) impact the validity of ACCESS for students with varying levels of prior knowledge or cognitive development?
3. How does student test anxiety or unfamiliarity with standardized testing affect score reliability compared to classroom performance?
4. How consistent are ACCESS scores across different administrations for the same student over time?

Literature Review

It is undeniable that high-stakes testing continues to serve as a central topic of vigorous scholarly debate among educators and researchers. In actuality, the proponents of standardized assessments (Allen 1995, Stotsky 2000, Allen 2001, Hanushek 2003, Hanushek & Raymond 2005, and Bowdon 2010) argue that these tests promote accountability and uphold consistent academic standards. However, a substantial body of evidence challenges the subtlety and grandeur of such tests at a larger scale. To address this gap in the literature, the current research seeks to primarily provide a systematic review on WIDA's ACCESS test validity and reliability between 2005 and 2025.

First of all, it is essential to discriminate between the reliability of test scores and the validity of interpretations or uses of the scores. To establish a thorough understanding of validity and reliability, it is of paramount importance to define these technical terms clearly. "Validity refers to the degree to which a specific interpretation or use of a test score is supported by the accumulated evidence (AERA, APA, & NCME, 2014; ETS, 2002). Validity is the central notion underlying the development, administration, and scoring of a test and the uses and interpretations of test scores (AERA et al., 2014. p. 11)." However, according to the Standards for Educational and Psychological Testing reliability is defined as "the consistency over replications of the testing procedure. Reliability/precision is high if the testing scores for each person are consistent over replications of the testing procedure and is low if the testing scores are not consistent over replications (AERA at al., 2014, p. 33)."

It must be noted that test validity can be influenced by numerous factors that go beyond individual student performance, including test-takers' anxiety (Al Fraidan, 2025), digital literacy, cultural background (Brookhart, 2012), prior instructional experiences (Cross, 2025), assessment design and scoring rubrics (Jonsson & Svingby, 2007), administration procedures (AERA, APA, & NCME, 2014), and classroom practices (Brookhart, 2012). In effect, student anxiety can negatively affect performance, though active-learning strategies may reduce it (Downing, 2020, p. 60). Digital literacy impacts outcomes, as students' proficiency affects resource use (Aydınlar et al., 2024, p. 38), and digital media can lower anxiety while improving language skills.

Cultural and social factors affect fairness in testing (Bennett, 2023, pp. 389–411), and students' prior learning influences test validity. Analytic rubrics give more reliable scores and feedback than holistic rubrics (Jonsson & Svingby, 2007, pp. 130–144), while standard test procedures ensure comparability (Berman, 2020, p. 22). Classroom practices, such as differentiation and self-management also support success (Smith, 2022, p. 34). McNeal (2016) found that "discrepancies in scores, particularly in the listening domain, suggest that ACCESS may not consistently measure students' language proficiency across all domains (p. 12)." Moreover, McNeal (2016) noted that educators may rely on composite scores for placement decisions, potentially overlooking individual strengths or weaknesses in particular domains. Waters (2021) further complicates the discussion by noting that "differences between paper-and-pencil and online test formats may impact the reliability of ACCESS scores, raising concerns for high-stakes decision-making (p. 4)." This example highlights the real-world implications of reliability concerns and reinforces the argument that multiple assessment measures are essential. Wolf (2008) argued that "current English language proficiency assessments, including ACCESS, often fail to fully capture the complex linguistic skills of ELLs (p. 360)." This limitation is particularly salient in content-based assessment items, where language proficiency may be conflated with prior content knowledge. Salazar (2022) found that such items "conflate language proficiency with content knowledge, limiting their construct validity (p. 78)." Wolf (2010) further emphasized that assessment tools may be "insufficiently sensitive to the diverse linguistic

backgrounds of students" (p. 530). For instance, a bilingual student may excel in conversational English yet struggle with formal academic language. ACCESS may fail to differentiate these competencies, leading to potential misplacement in language support programs.

In the same vein, Stefanakis (2010) noted that "standardized tests like ACCESS may not account for the cultural and linguistic diversity of ELL students, potentially leading to misidentification of their academic abilities" (p. 5). Cultural familiarity with test content can directly influence writing production; for example, a student from Morocco may face limitations in writing about Thanksgiving or Halloween, as these holidays are not part of Moroccan culture. Consequently, the student might lack the background knowledge needed to fully understand or accurately discuss these topics. Similarly, Salinas (2022) notes that standardized tests lack cultural relevance and equitable representation, disadvantaging students from underrepresented groups.

Espinosa (2010) highlighted that "many assessments, including ACCESS, fail to adequately capture the cognitive and linguistic development of young learners (p. 7)." Furthermore, Waters (2021) cautioned that "ACCESS scores should not be the sole determinant for program placement or instructional decisions, as reliability and validity concerns remain (p. 6)." Additionally, Wolf (2008) similarly warned that "over-reliance on a single standardized test can lead to misinterpretation of students' abilities (p. 362)", while McNeal (2016) emphasized the need for "triangulated evidence, highlighting correlations between ACCESS and other language assessments that demonstrate inconsistencies (p. 15)." Wolf (2010) concluded that "valid and equitable assessment of ELLs requires ongoing refinement of instruments like ACCESS, coupled with professional judgment and multiple indicators of language proficiency (p. 540)." Villegas (2022) echoed this perspective, noting that the limitations of ACCESS "are not reasons to abandon it, but to use it thoughtfully and in conjunction with other measures (para. 4)."

Furthermore, a mixed-methods study conducted by Waters (2020) at VCU examined teachers' perceptions of the WIDA's ACCESS test. The study found that educators identified several threats to the test's validity, including issues with technology associated with the online version of the test, the lapse of time between test administration and the receipt of scores, and questions regarding student motivation and test-taking effort. These factors signify that the test may not consistently produce accurate or fair results for multilingual learners. Moreover, the study indicated that the test has unintended consequences, such as negative emotional impacts for teachers and students and a loss of instructional time. Waters (2020) emphasized the importance of considering teacher input in test evaluations and suggested the use of multiple measures in high-stakes decision-making for English learners.

From a policy standpoint, external reports and technical reviews have critiqued the accessibility features of the ACCESS test. The 2019-2020 Accessibility and Accommodations Supplement emphasize the importance of individualized accommodations for English language learners to ensure valid assessment results. However, the document also notes that these accommodations and embedded digital tools may produce inconsistent outcomes for multilingual learners, indicating potential challenges in achieving equitable assessment practices (State Superintendent of Education, 2020). According to Belinder et al., "our research informs us that high-stakes testing is hurting students, teachers, and schools. It is putting the nation at risk. By restricting the education of our young people and substituting for it training for performing well on high-stakes examinations, we are turning America into a nation of test-takers, abandoning our heritage as a nation of thinkers, dreamers, and doers (2007, p. 3)." These authors warn that focusing too much on high-stakes testing harms real learning, turning students into test-takers instead of creative thinkers and problem-solvers. Other critics expressed the need of integrating performance-based and classroom-centered measures. For example, Frederiksen argued that "traditional multiple-choice tests distort teaching and fail to reflect students' authentic abilities, advocating for tasks mirroring real-world contexts (Frederiksen, 1984, p. 23)." In this regard, Ly suggests alternative assessments, such as portfolios, performance tasks, and dynamic assessment, which offer a more authentic view of ELs' language development than standardized tests alone (Ly, 2024, p. 178; Jalilzadeh, 2023, p. 7). In effect, through emphasizing both cognitive skills and sociocultural aspects of language acquisition, these assessment approaches not only supplement ACCESS scores, but also promote inclusive and formative teaching practices (Avdiu & Ahmedi, 2024, p. 15). Additionally, Hadjioannou (2017) explains that "the problems for ELs will persist unless significant changes are made, and: The time to address assessment concerns for ELs in US schools is now. Without a widespread

effort to reevaluate the academic support and corresponding negative consequences for this failure among ELs will have significant consequences for maintaining and sustaining democratic and diversified schools (p. 121)."

One should know that he deployment of this extensive literature review seeks to extend the scope of this research, positioning it at the forefront of academic inquiry and providing a comprehensive understanding of the dynamic interplays and thematic evolutions within the field of English as a Second Language (ESL) in general and Bilingual Education in particular. Synthesizing the literature reveals several intersecting themes, namely persistent reliability issues influenced by test format and administration; limited construct validity due to overlap between content knowledge and language proficiency; performance shaped by cultural, linguistic, and developmental factors; and emerging equity concerns from technological and automated scoring systems. These findings raise a growing concern regarding the variability in ACCESS scores, which higlights the importance of approaching standardized test data cautiously, particularly when making high-stakes decisions.

ACCESS Test Preliminaries

Figure 1

Language Proficiency levels

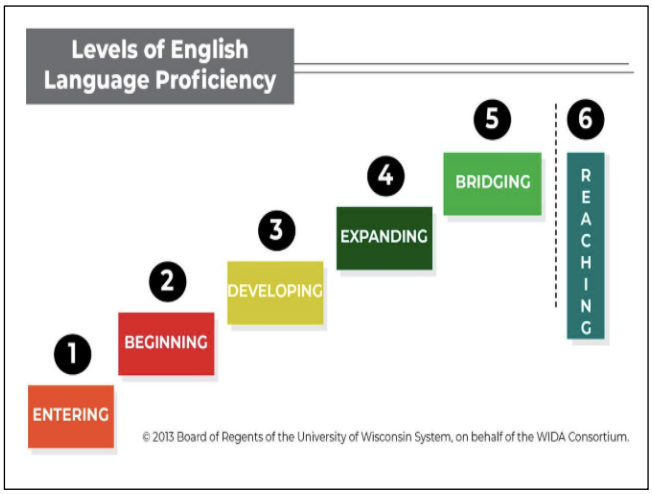
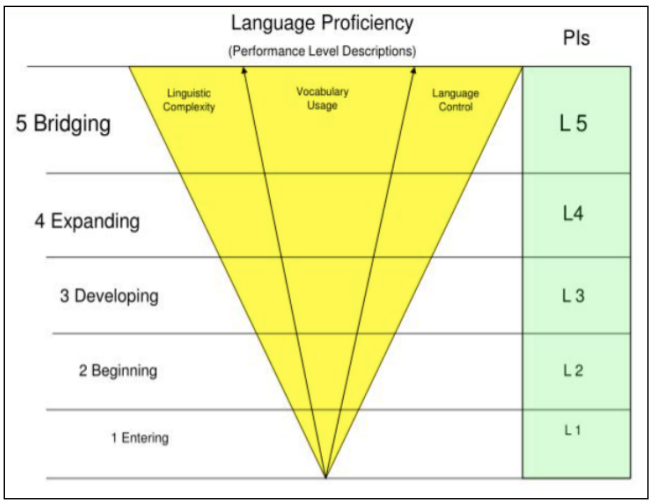


Figure 2

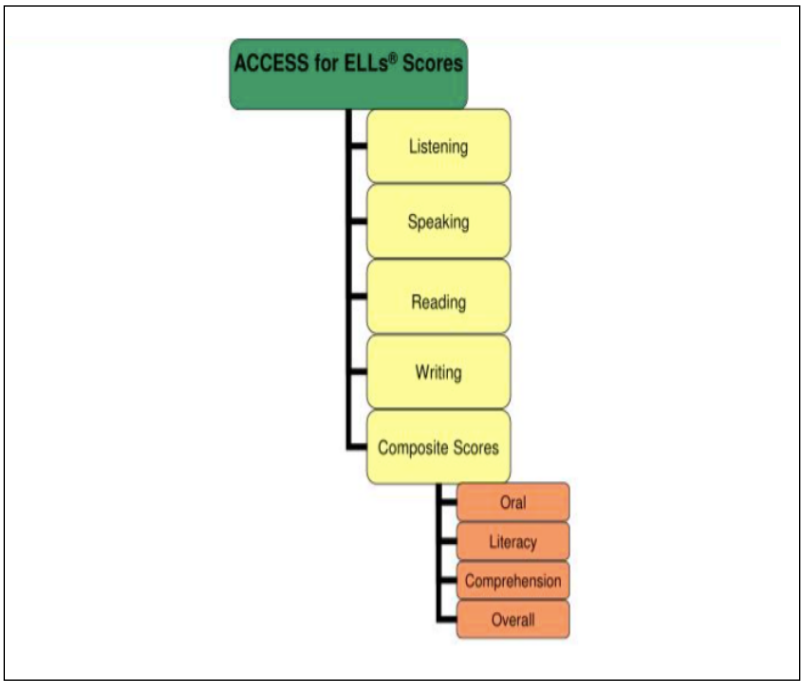
Performance Level Descriptors



The charts illustrate the six proficiency levels of the ACCESS test, ranging from Level 1 (Entering), representing minimal English ability, to Level 6 (Reaching), indicating advanced English proficiency. Each level reflects increasing language skills in listening, speaking, reading, and writing, providing a clear framework for assessing English learners' proficiency over time.

Figure 3

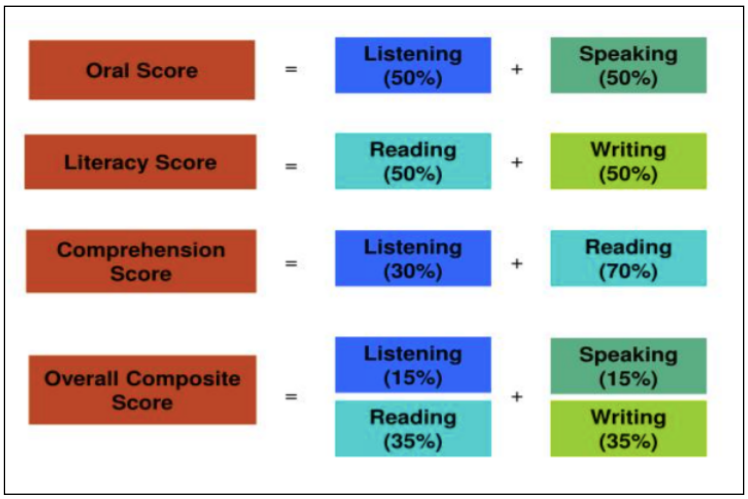
Main ACCESS Language Components



The chart portrays key components assessed by the ACCESS test for ELs. It is divided into four main domains: Listening, Speaking, Reading, and Writing. Each domain may include subskills, such as oral, literacy, and comprehension showing how students' language proficiency is measured across multiple modalities. The chart visually represents these areas to highlight the comprehensive nature of the ACCESS test.

Figure 4

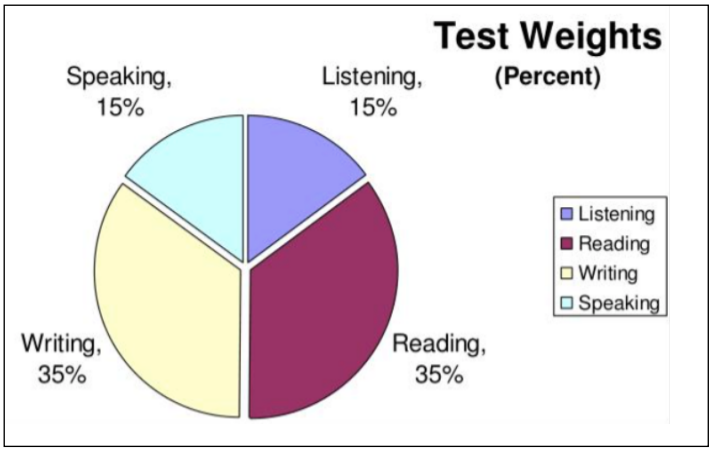
Composite Scores



The Oral score combines the Speaking (50%) and Listening (50%) domains and reflects a student's ability to understand spoken English and communicate orally in both academic and social contexts. The Literacy score combines Reading (50%) and Writing (50%), assessing a student's ability to comprehend written text and express ideas in writing, providing a measure of academic literacy. The Comprehension score combines Listening (30%) and Reading (70%), assessing a student's overall ability to understand and interpret both spoken and written English in academic and social contexts. The Overall Composite score combines all domains. Scores range from 1.0 to 6.0, reflecting WIDA's proficiency levels in order to plan instruction and guide students' progress accordingly.

Figure 5

Test Weights as Percentage of Overall Composite Score



As shown in the pie chart, the ACCESS test measures English in four domains: Listening (15%), Speaking (15%), Reading (35%), and Writing (35%), together forming the overall language proficiency score.

Figure 6

ACCESS Scale Scores (K–12)

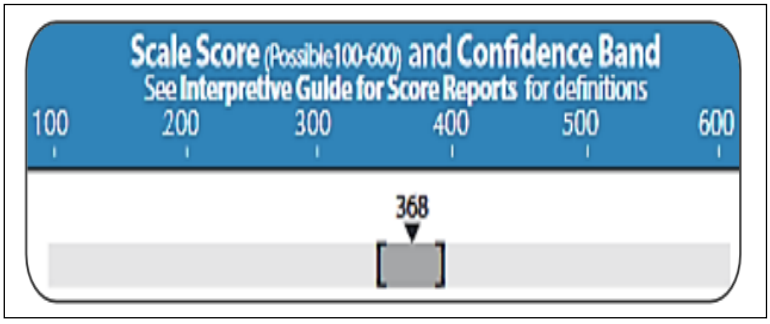
Elementary	
Grade	Scale Score
K	319
1	339
2	353
3	365
4	376
5	386

Middle & High	
Grade	Scale Score
6	393
7	400
8	406
9	412
10	418
11	423
12	428

These charts show scale scores for each grade level (K–12). Scores increase with grade level, reflecting the rising complexity of language tasks and higher proficiency expectations.

Figure 7

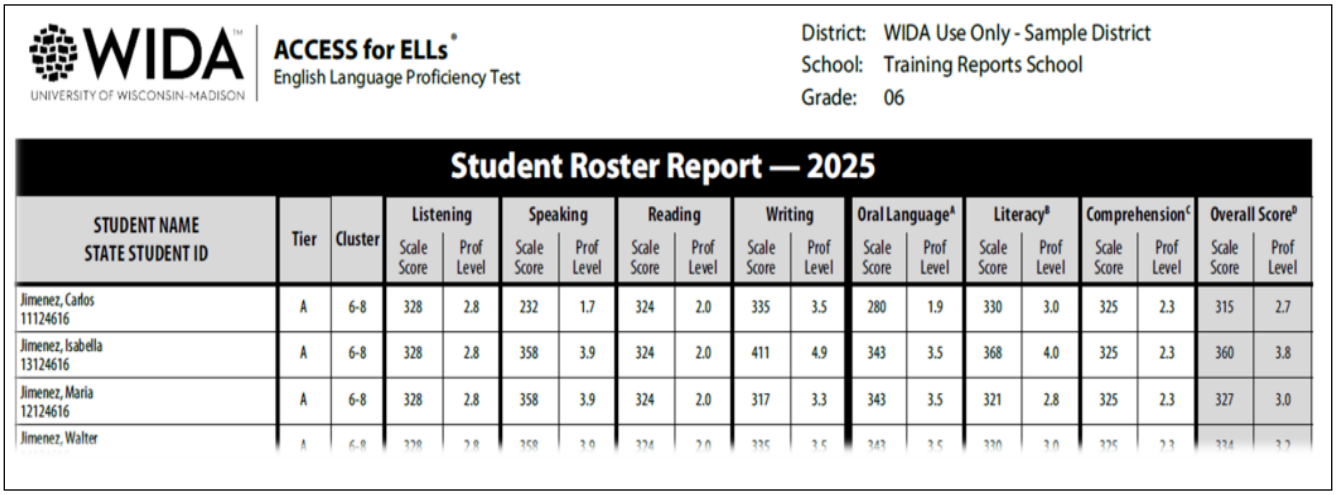
Confidence Bands (WIDA 2025)



Confidence bands on the Individual Student Report show the Standard Error of Measurement (SEM) around a student's scale score. The shaded box beneath the score indicates the range the student might score if the test was taken multiple times. These bands remind us that a scale score is an estimate influenced by testing conditions—such as health, rest, or distractions during the session (WIDA, 2025, p. 6).

Figure 8

Student Roster Report (WIDA 2025)



WIDA UNIVERSITY OF WISCONSIN-MADISON | ACCESS for ELLs® English Language Proficiency Test

District: WIDA Use Only - Sample District
School: Training Reports School
Grade: 06

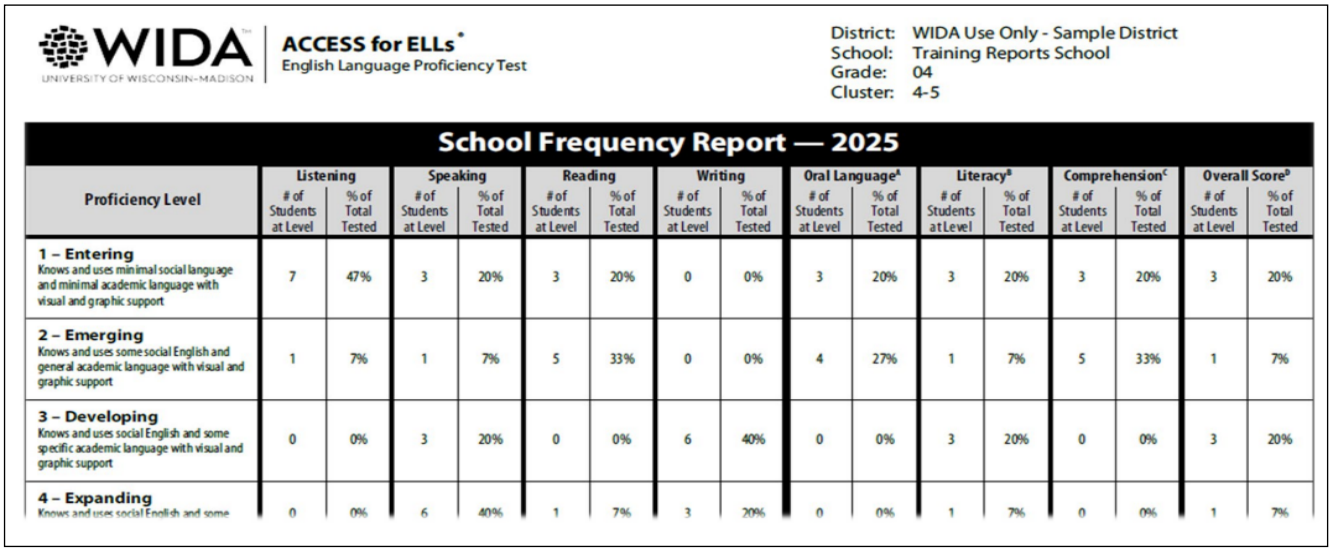
Student Roster Report — 2025

STUDENT NAME / STATE STUDENT ID	Tier	Cluster	Listening		Speaking		Reading		Writing		Oral Language[A]		Literacy[B]		Comprehension[C]		Overall Score[D]	
			Scale Score	Prof Level	Scale Score	Prof Level	Scale Score	Prof Level	Scale Score	Prof Level	Scale Score	Prof Level	Scale Score	Prof Level	Scale Score	Prof Level	Scale Score	Prof Level
Jimenez, Carlos 11124616	A	6-8	328	2.8	232	1.7	324	2.0	335	3.5	280	1.9	330	3.0	325	2.3	315	2.7
Jimenez, Isabella 13124616	A	6-8	328	2.8	358	3.9	324	2.0	411	4.9	343	3.5	368	4.0	325	2.3	360	3.8
Jimenez, Maria 12124616	A	6-8	328	2.8	358	3.9	324	2.0	317	3.3	343	3.5	321	2.8	325	2.3	327	3.0
Jimenez, Walter	A	6-8	328	2.8	358	3.9	324	2.0	335	3.5	343	3.5	330	3.0	325	2.3	334	3.2

The Student Roster Report provides scale scores and proficiency levels by language domain and overall performance for a specific school and grade. It helps educators identify performance trends, verify scores, group students for targeted instruction, and guide school or district improvement efforts (WIDA 2025, p. 10).

Figure 9

Frequency Reports (WIDA 2025)



WIDA UNIVERSITY OF WISCONSIN-MADISON | ACCESS for ELLs® English Language Proficiency Test

District: WIDA Use Only - Sample District
School: Training Reports School
Grade: 04
Cluster: 4-5

School Frequency Report — 2025

Proficiency Level	Listening		Speaking		Reading		Writing		Oral Language[A]		Literacy[B]		Comprehension[C]		Overall Score[D]	
	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested	# of Students at Level	% of Total Tested
1 – Entering Knows and uses minimal social language and minimal academic language with visual and graphic support	7	47%	3	20%	3	20%	0	0%	3	20%	3	20%	3	20%	3	20%
2 – Emerging Knows and uses some social English and general academic language with visual and graphic support	1	7%	1	7%	5	33%	0	0%	4	27%	1	7%	5	33%	1	7%
3 – Developing Knows and uses social English and some specific academic language with visual and graphic support	0	0%	3	20%	0	0%	6	40%	0	0%	3	20%	0	0%	3	20%
4 – Expanding Knows and uses social English and some	0	0%	6	40%	1	7%	3	20%	0	0%	1	7%	0	0%	1	7%

Frequency reports provide an overview of student performance by showing the number and percentage of students at each proficiency level, along with the highest and lowest scores in the four language domains. They help administrators and policymakers plan English language support and compare the progress of English learners with their peers (WIDA 2025, p. 11).

Table 1

Comparison between TOEFL, ACCESS, and ACTFL

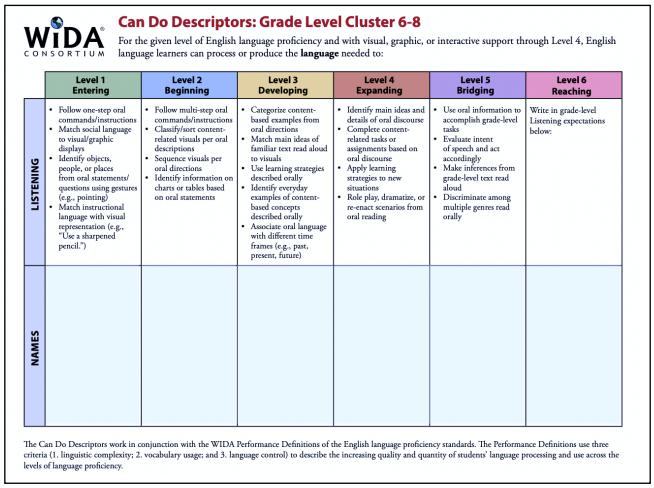
Test Type	TOEFL (Test of English as a Foreign Language)	ACCESS for ELLs (Assessing Comprehension and Communication in English State-to-State)	ACTFL Assessment of Performance toward Proficiency in Languages® (AAPPL)
Purpose	Measures English proficiency for academic admission (college/university).	Measures English language development and proficiency for K–12 students in U.S. schools.	Measures proficiency/performance in target languages (including English) across modes of communication; used for placement, certification, proficiency benchmarking. (ACTFL)
Target Group	Non-native English speakers applying to higher education.	English Language Learners (ELLs) enrolled in U.S. public schools (K–12).	Learners (K-12 and postsecondary) of world languages (including English) or professional/educator certification contexts. (Default)
Developer	Educational Testing Service (ETS) (ETS, 2025)	WIDA Consortium (WIDA, 2023)	ACTFL, through its assessments (via Language Testing International (LTI) as administrator) (ACTFL)
Use Context	Used by colleges/universities to assess readiness for academic study.	Used by schools to monitor progress, plan instruction, and determine program eligibility.	Used by educational institutions, school systems, agencies for placement, program evaluation, certification, and proficiency benchmark. (ACTFL)
Test Content	Focuses on academic English — reading, listening, speaking, writing in academic contexts.	Aligns with academic standards and language development levels; reflects classroom language in content areas.	Assesses language proficiency and performance across listening, reading, speaking, writing modes, often using real-world tasks and aligned with the ACTFL Proficiency Guidelines. (ACTFL)
Scoring and Outcomes	Produces a single overall score (0–120) for proficiency.	Produces proficiency levels (1–6) showing stages of English development for instructional planning.	Provides proficiency ratings (e.g., Novice, Intermediate, Advanced, Superior) or performance levels; used for placement, credit, certification. (CLA Language Testing Program)
Administration	Taken individually, usually online or at test centers.	Administered annually in schools to all identified ELL students (group setting).	Administered via computer or interview formats (live or computer-based) depending on the sub-assessment; K-12 or institutional settings. (ACTFL)
High-Stakes Use	High-stakes for admission and scholarships.	High-stakes for educational placement, accountability, and program evaluation.	High-stakes for language certification, placement, bilingual program eligibility, educator licensure or program accreditation.
Test Length & Format	About 2–3 hours, adaptive online test.	Varies by grade cluster; typically 2–3 hours across several sessions.	Varies widely by test type (e.g., interview for speaking proficiency, computer adaptive for listening/reading); formats may include one-on-one oral, computer adaptive modules, writing tasks.

It is striking that TOEFL, ACCESS, and ACTFL differ in focus and use. TOEFL measures academic English for college admission with a single score. ACCESS tracks K–12 English learners' development with proficiency levels for instruction and placement, while ACTFL assesses practical language skills across levels for placement, certification, and program evaluation.

WIDA's Can Do Descriptors (6-8)

Figure 10

Listening



Can Do Descriptors: Grade Level Cluster 6-8

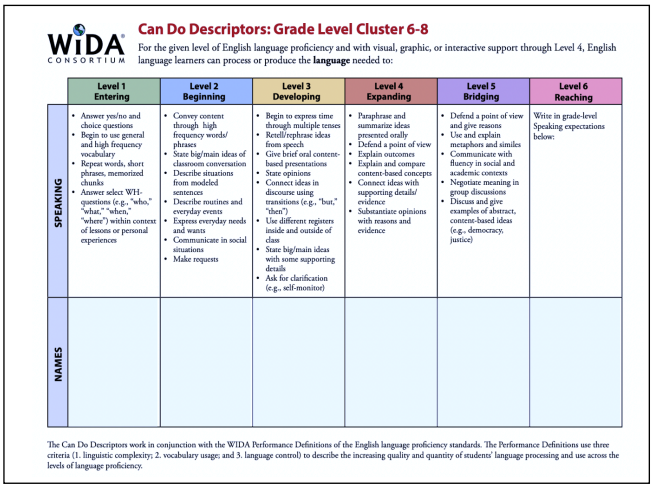
For the given level of English language proficiency and with visual, graphic, or interactive support through Level 4, English language learners can process or produce the **language** needed to:

	Level 1 Entering	Level 2 Beginning	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 Reaching
LISTENING	• Follow one-step oral commands/instructions • Match social language to visual/graphic displays • Identify objects, people, or places from oral statements/ questions using gestures (e.g., pointing) • Match instructional language with visual representation (e.g., "Use a sharpened pencil.")	• Follow multi-step oral commands/instructions • Classify/sort content-related visuals per oral descriptions • Sequence visuals per oral directions • Identify information on charts or tables based on oral statements	• Categorize content-based examples from oral directions • Match main ideas of familiar text read aloud to visuals • Use learning strategies described orally • Identify everyday examples of content-based concepts described orally • Associate oral language with different time frames (e.g., past, present, future)	• Identify main ideas and details of oral discourse • Complete content-related tasks or assignments based on oral discourse • Apply learning strategies to new situations • Role play, dramatize, or re-enact scenarios from oral reading	• Use oral information to accomplish grade-level tasks • Evaluate intent of speech and act accordingly • Make inferences from grade-level text read aloud • Discriminate among multiple genres read orally	Write in grade-level Listening expectations below:
NAMES						

The Can Do Descriptors work in conjunction with the WIDA Performance Definitions of the English language proficiency standards. The Performance Definitions use three criteria (1. linguistic complexity; 2. vocabulary usage; and 3. language control) to describe the increasing quality and quantity of students' language processing and use across the levels of language proficiency.

Figure 11

Speaking



Can Do Descriptors: Grade Level Cluster 6-8

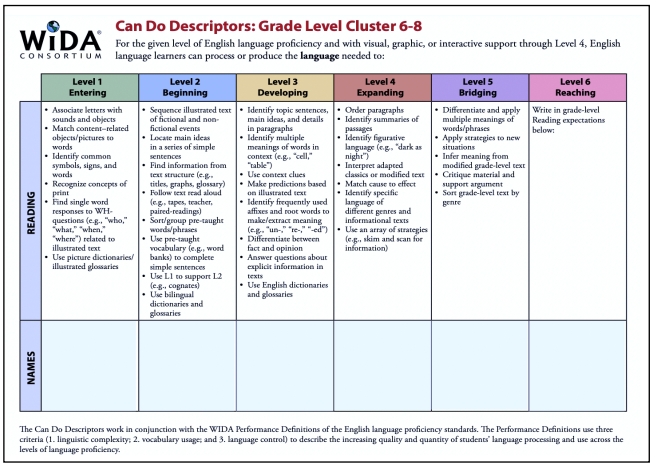
For the given level of English language proficiency and with visual, graphic, or interactive support through Level 4, English language learners can process or produce the **language** needed to:

	Level 1 Entering	Level 2 Beginning	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 Reaching
SPEAKING	• Answer yes/no and choice questions • Begin to use general and high frequency vocabulary • Repeat words, short phrases, memorized chunks • Answer select WH-questions (e.g., "who," "what," "when," "where") within context of lessons or personal experiences	• Convey content through high frequency words/ phrases • State big/main ideas of classroom conversation • Describe situations from modeled sentences • Describe routines and everyday events • Express everyday needs and wants • Communicate in social situations • Make requests	• Begin to express time through multiple tenses • Retell/rephrase ideas from speech • Give brief oral content-based presentations • State opinions • Connect ideas in discourse using transitions (e.g., "but," "then") • Use different registers inside and outside of class • State big/main ideas with some supporting details • Ask for clarification (e.g., self-monitor)	• Paraphrase and summarize ideas presented orally • Defend a point of view • Explain outcomes • Explain and compare content-based concepts • Connect ideas with supporting details/ evidence • Substantiate opinions with reasons and evidence	• Defend a point of view and give reasons • Use and explain metaphors and similes • Communicate with fluency in social and academic contexts • Negotiate meaning in group discussions • Discuss and give examples of abstract, content-based ideas (e.g., democracy, justice)	Write in grade-level Speaking expectations below:
NAMES						

The Can Do Descriptors work in conjunction with the WIDA Performance Definitions of the English language proficiency standards. The Performance Definitions use three criteria (1. linguistic complexity; 2. vocabulary usage; and 3. language control) to describe the increasing quality and quantity of students' language processing and use across the levels of language proficiency.

Figure 10

Reading



Can Do Descriptors: Grade Level Cluster 6-8

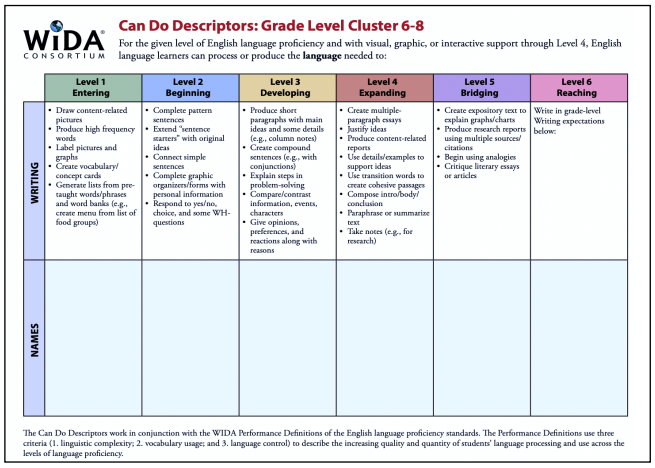
For the given level of English language proficiency and with visual, graphic, or interactive support through Level 4, English language learners can process or produce the **language** needed to:

	Level 1 Entering	Level 2 Beginning	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 Reaching
READING	• Associate letters with sounds and objects • Match content-related objects/pictures to words • Identify common symbols, signs, and words • Recognize concepts of print • Find single word responses to WH-questions (e.g., "who," "what," "when," "where") related to illustrated text • Use picture dictionaries/ illustrated glossaries	• Sequence illustrated text of fictional and non-fictional events • Locate main ideas in a series of simple sentences • Find information from text structure (e.g., titles, graphs, glossary) • Follow text read aloud (e.g., tapes, teacher, paired-readings) • Sort/group pre-taught words/phrases • Use pre-taught vocabulary (e.g., word banks) to complete simple sentences • Use L1 to support L2 (e.g., cognates) • Use bilingual dictionaries and glossaries	• Identify topic sentences, main ideas, and details in paragraphs • Identify multiple meanings of words in context (e.g., "cell," "table") • Use context clues • Make predictions based on illustrated text • Identify frequently used affixes and root words to make/extract meaning (e.g., "un-," "re-," "-ed") • Differentiate between fact and opinion • Answer questions about explicit information in texts • Use English dictionaries and glossaries	• Order paragraphs • Identify summaries of passages • Identify figurative language (e.g., "dark as night") • Interpret adapted classics or modified text • Match cause to effect • Identify specific language of different genres and informational texts • Use an array of strategies (e.g., skim and scan for information)	• Differentiate and apply multiple meanings of words/phrases • Apply strategies to new situations • Infer meaning from modified grade-level text • Critique material and support argument • Sort grade-level text by genre	Write in grade-level Reading expectations below:
NAMES						

The Can Do Descriptors work in conjunction with the WIDA Performance Definitions of the English language proficiency standards. The Performance Definitions use three criteria (1. linguistic complexity; 2. vocabulary usage; and 3. language control) to describe the increasing quality and quantity of students' language processing and use across the levels of language proficiency.

Figure 11

Writing



Can Do Descriptors: Grade Level Cluster 6-8

For the given level of English language proficiency and with visual, graphic, or interactive support through Level 4, English language learners can process or produce the **language** needed to:

	Level 1 Entering	Level 2 Beginning	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 Reaching
WRITING	• Draw content-related pictures • Produce high frequency words • Label pictures and graphs • Create vocabulary/ concept cards • Generate lists from pre-taught words/phrases and word banks (e.g., create menu from list of food groups)	• Complete pattern sentences • Extend "sentence starters" with original ideas • Connect simple sentences • Complete graphic organizers/forms with personal information • Respond to yes/no, choice, and some WH-questions	• Produce short paragraphs with main ideas and some details (e.g., column notes) • Create compound sentences (e.g., with conjunctions) • Explain steps in problem-solving • Compare/contrast information, events, characters • Give opinions, preferences, and reactions along with reasons	• Create multiple-paragraph essays • Justify ideas • Produce content-related reports • Use details/examples to support ideas • Use transition words to create cohesive passages • Compose intro/body/ conclusion • Paraphrase or summarize text • Take notes (e.g., for research)	• Create expository text to explain graphs/charts • Produce research reports using multiple sources/ citations • Begin using analogies • Critique literary essays or articles	Write in grade-level Writing expectations below:
NAMES						

The Can Do Descriptors work in conjunction with the WIDA Performance Definitions of the English language proficiency standards. The Performance Definitions use three criteria (1. linguistic complexity; 2. vocabulary usage; and 3. language control) to describe the increasing quality and quantity of students' language processing and use across the levels of language proficiency.

These charts display WIDA's Can Do Descriptors. Such descriptors describe what English learners can typically do at various stages (Entering-Reaching) of language proficiency across the four domains: listening, speaking, reading, and writing. They are organized in clusters to provide functional and real-world examples of language use. For illustration, the Grades 6–8 Can Do Descriptors are provided above as an example to help readers become more familiar with how these descriptors guide instruction. Other clusters are k, 1-2, 3-5, and 9-12. They are also organized in the same fashion.

Methodology

This study employs a dual-methodological framework integrating both bibliometric and content analysis methods to rigorously examine the validity and reliability of WIDA's ACCESS test. In doing so, the current research provides a comprehensive and nuanced understanding of the assessment, aligning with the study's multifaceted research questions and its aim to interrogate complex dimensions of language testing.

Bibliometric Analysis

Bibliometric analysis is a systematic and quantitative method for studying scholarly literature, focusing on patterns of publications, citations, and authorship. It allows researchers to uncover key contributions, track emerging research trends, and map collaboration networks within a given discipline. According to Celestino et al. (2024), bibliometric

analysis is a "methodology for retrospective, quantitative, statistical, and descriptive study of the scientific literature," noting its utility in identifying prolific authors, influential publications, and emerging research themes (p.14).

Content Analysis

Content analysis is a research method used to systematically analyze and interpret the content of various forms of communication, such as texts, images, or audio. It involves categorizing and coding content to identify patterns, themes, or biases, allowing researchers to make inferences about the messages conveyed. As Krippendorff defines it, "content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (1980, p. 21).

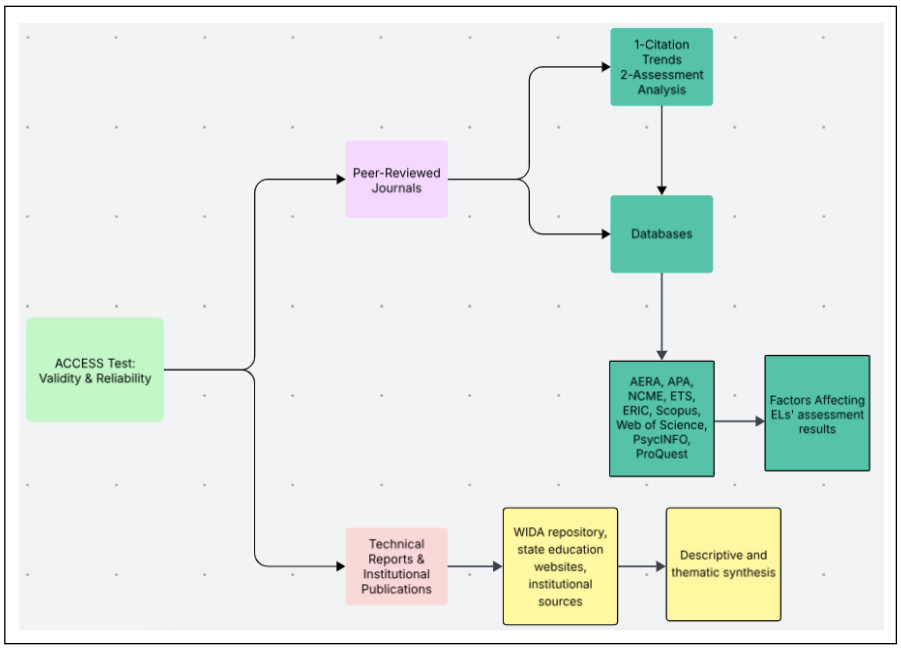
Data Collection Methods & Procedure

This study employed a rigorous and systematic data collection procedure centered on an extensive review of peer-reviewed journal articles, scholarly studies, technical reports, and institutional publications that examined the validity and reliability of the WIDA ACCESS for ELLs assessment. To ensure comprehensive coverage, searches were conducted across major academic databases, including ERIC, Scopus, Web of Science, PsycINFO, and ProQuest Dissertations & Theses, as well as targeted searches within WIDA's official repository, state education department websites, and other relevant institutional sources.

Titles and abstracts were screened for relevance. Key data on research design, participants, analytical methods, and reliability and validity findings were systematically extracted, and technical reports and institutional publications were appraised for methodological rigor. The data was synthesized descriptively and thematically to identify patterns, gaps, and evidence regarding the ACCESS test's validity and reliability, psychometric properties in order to ensure a robust and replicable review process.

Figure 13

Data Collection Plan



Data Analysis & Discussion

This section provides a systematic review of over 30 scholarly articles, research studies, and academic papers, critically evaluating the theoretical foundation of the ACCESS test in order to identify patterns, trends, and gaps in the existing literature. This process enabled a thorough examination of the ACCESS test's psychometric properties, particularly with respect to its construct validity, reliability, and overall effectiveness as a measure of English language proficiency, while shedding light upon the inherent challenges and limitations tackled in previously published studies.

Table 2

Publication Trends from 2005 to 2025

Year	Author(s) / Organization	Title	Focus Area	Key Findings
2005	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 100–2004–2005 Administration	Technical Report	Established the foundational reliability and validity of ACCESS for ELLs for K–12 students.
2006	Bauman et al.	ACCESS for ELLs® Test Design	Test Design	Discussed the integration of social instructional language and academic English in test design.
2007	WIDA Consortium	English Language Development Standards	Standards Framework	Introduced the WIDA ELD Standards, forming the basis for ACCESS assessments.
2008	Wolf et al.	Validity Issues in Assessing English Language Learners' Language Proficiency	Validity	Examined construct validity concerns in ELL assessments, including ACCESS.
2009	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 200–2008–2009 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2010	Wolf et al.	Improving the Validity of English Language Learner Assessments	Validity	Proposed strategies to enhance the validity of ELL assessments like ACCESS.
2011	Short et al.	Research on Academic Literacy Development in Sheltered Instruction Classrooms	Literacy Development	Investigated the impact of sheltered instruction on ELL literacy, informing ACCESS test design.
2012	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 300–2011–2012 Administration	Technical Report	Provided insights into the 2011–2012 administration of ACCESS for ELLs.
2013	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 400–2012–2013 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2015	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 500–2014–2015 Administration	Technical Report	Provided insights into the 2014–2015 administration of ACCESS for ELLs.
2016	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 600–2015–2016 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2016	Callahan et al.	Equitable Access for Secondary English Learner Students	Equity	Analyzed high school course-taking patterns to assess equity in ELL education.
2016	McNeal	Correlating English Language Learner CRCT Scores on the ACCESS for ELLs Assessment	Assessment Comparison	Compared ACCESS scores with other assessments for ELLs.
2017	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 700–2016–2017 Administration	Technical Report	Provided insights into the 2016–2017 administration of ACCESS for ELLs.

Year	Author(s) / Organization	Title	Focus Area	Key Findings
2018	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 800–2017–2018 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2019	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 900–2018–2019 Administration	Technical Report	Provided insights into the 2018–2019 administration of ACCESS for ELLs.
2020	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1000–2019–2020 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2021	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1100–2020–2021 Administration	Technical Report	Provided insights into the 2020–2021 administration of ACCESS for ELLs.
2021	Waters	Considerations for Using ACCESS Test Scores in Decision Making	Policy & Practice	Discussed the implications of ACCESS scores in educational decision-making.
2022	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1200–2021–2022 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2022	Villegas	Long-standing Limitations of English Learner Academic Assessment Data Persist	Policy & Equity	Discussed limitations in ELL assessment data and implications.
2023	Schechter	Exploration of ELL Student WIDA Access Growth Grades 1–5	Student Growth	Investigated ELL student growth using WIDA ACCESS scores.
2023	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1300–2022–2023 Administration	Technical Report	Provided insights into the 2022–2023 administration of ACCESS for ELLs.
2024	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1400–2023–2024 Administration	Technical Report	Reported on the administration and performance metrics of ACCESS for ELLs.
2024	Patterson	Are WIDA Test Results Appropriately Reflecting Teacher Experiences?	Teacher Perspectives	Examined ESOL teachers' experiences with WIDA assessments.
2024	Gao et al.	Automatic Assessment of Text-Based Responses in Post-Secondary Education: A Systematic Review	Technology & Assessment	Reviewed automated assessment systems in education, with implications for ELLs.
2024	Huang et al.	Enhancing Essay Scoring with Adversarial Weights Perturbation and Metric-specific Attention Pooling	Technology & Assessment	Explored methods to improve automated essay scoring for ELLs.
2025	WIDA Consortium	Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1500–2024–2025 Administration	Technical Report	Provided insights into the 2024–2025 administration of ACCESS for ELLs.

Phase 1: Early Development and Initial Validation (2005–2010)

During the initial phase, the focus was on the creation and pilot testing of the ACCESS for ELLs assessment. Efforts concentrated on ensuring that the assessment aligned with WIDA English Language Development (ELD) standards, while establishing foundational content and construct validity. Early technical reports documented basic psychometric properties and reliability metrics, though research outputs were limited, consisting mainly of internal WIDA reports and small-scale studies. Foundational scoring rubrics were developed, and initial reliability analyses emphasized internal consistency and inter-rater agreement. This period laid the groundwork for subsequent empirical and psychometric expansion.

Phase 2: Expansion and Psychometric Strengthening (2011–2018)

The second phase marked an expansion in both the scope of research and methodological sophistication. Publications increased, including longitudinal and district-level studies, reflecting a broader empirical base. Statistical analyses became more advanced, incorporating Cronbach's alpha, inter-rater reliability, and Rasch modeling to strengthen evidence for reliability

Phase 3: Refinement, Technological Integration, and Advanced Validation (2019–2025)

The most recent phase focused on refinement, technological integration, and advanced psychometric validation. Emphasis shifted to digital administration and adaptation for online testing, alongside the application of multi-faceted Rasch models and other sophisticated psychometric techniques. There is a growing number of publicly available studies validated the writing and speaking scales and newly developed scoring rubrics. Annual technical reports became highly detailed, documenting analyses across paper and digital formats.

4.2 What is the Rasch model used for in WIDA's ACCESS test?

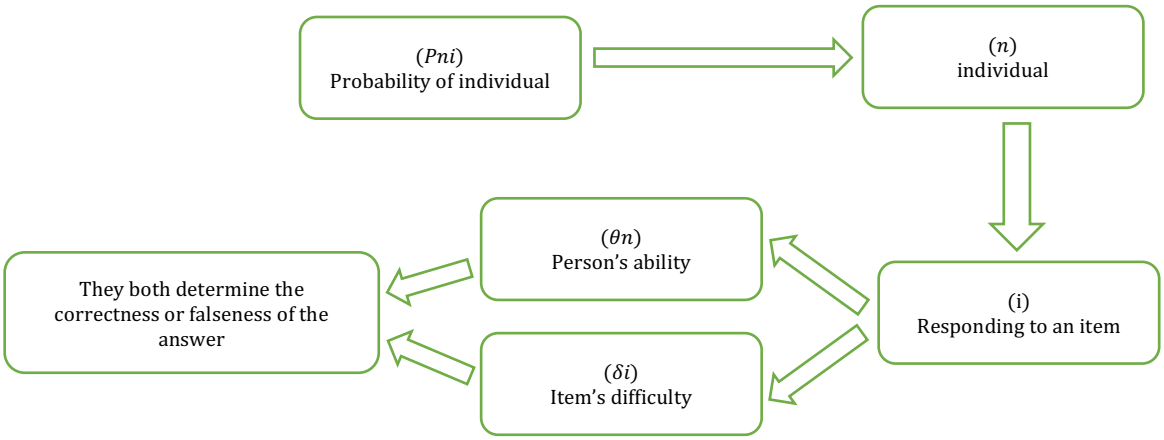
Rasch analysis (Rasch, 1960) is an Item Response Theory (IRT)-based method for test validation that provides an alternative to classical approaches. It assesses the psychometric properties of test items, informs questionnaire design, and can produce interval-level measures for latent traits (Liu & Boone, 2006). Unlike classical methods that analyze observed data, Rasch analysis uses a logistic model to estimate the probability (Pni) of an individual (n) responding to item (i) based on the difference between the person's ability (θn) and the item's difficulty (δi).

$$Pni = e\,(\theta n - \delta i)\,(1 + e\,(\theta n - \delta i)/).$$

Representation of the Rasch Model

Figure 14

Rash Model Representation



The Rasch model assumes a functional link between an item response and the probability of a student selecting it, represented by the item characteristic curve. In polytomous items, each response has its own curve, ordered by difficulty (Masters, 1982; Bond & Fox, 2015).

More importantly, the Rasch model converts ordinal data into interval scales via a logistic function, enabling evaluation of item properties such as unidimensionality, fit, difficulty, and reliability, including separation indices (Boone et al., 2010; Liu & Boone, 2006). For polytomous items, response categories must follow a logically increasing difficulty order. Overall, this ensures the data meet Rasch model requirements and supports the quality of measurement.

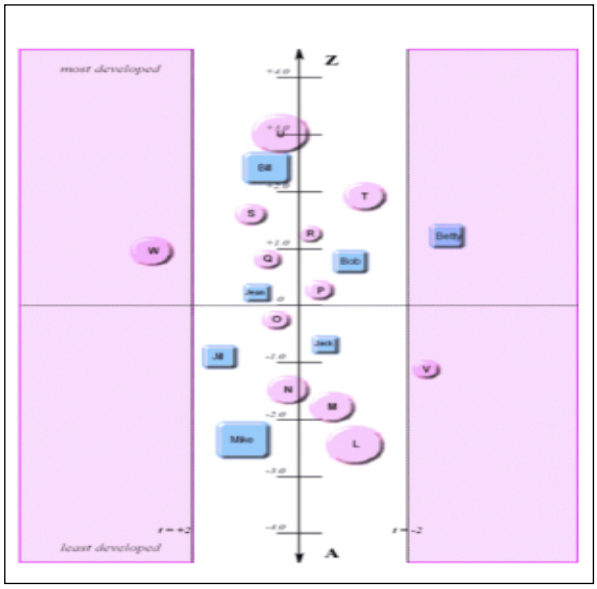
Applied Rasch Analysis

Calculate each person's percent correct, convert to odds p/(1−p)p/(1-p)p/(1−p), and log-transform to estimate ability (blue squares). Compute item difficulty similarly (pink circles). Plot both against the ideal Rasch model (vertical line, mean logit = 0). Symbol size reflects precision; error increases at extremes.

Assess fit: points outside the acceptable range (e.g., −2 to 2) indicate misfit. Revise items as needed. Once data align, total scores can be used as valid measures (Bond & Fox 2007).

Figure 15

Rasch Model Analysis



4.2.3 Mathematical Representation of the Rash Model

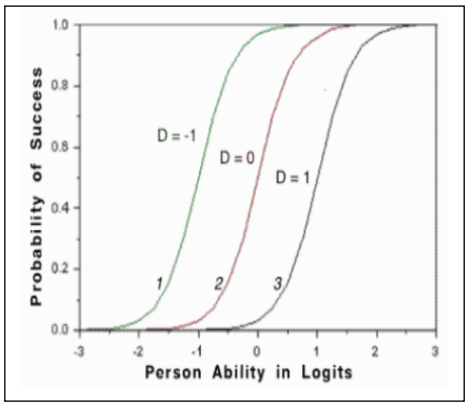
After estimating a person's ability (Bn)and an item's difficulty (Di), the likelihood of a correct response can be represented using Equation 1. This equation shows that the probability of subject n answering item i correctly is determined by a logistic function based on the difference between the person's ability and the item's difficulty (Bond & Fox 2007):

$$Pni(x=1)=f(Bn-Di) = e(Bn-Di)/1+e(Bn-Di), \text{ where } x=1 \text{ is correct \& } x=0 \text{ is incorrect } (1)$$

Model Fit

Figure 16

Model Fit



Item fit is assessed by comparing observed responses to model predictions using infit/outfit statistics and chi-square-based residuals, with standardized t-scores between -2 and +2 indicating acceptable fit; Wald and likelihood ratio tests identify misfitting items and evaluate overall model fit (Bond & Fox 2007).

Implications of the Rasch Model in WIDA's ACCESS Test

Rasch modeling, widely used in large-scale assessments like WIDA's ACCESS for ELLs, has notable strengths, but also limitations. Goldstein (1979, pp. 207–229) argued that the model's assumption of equal item discrimination may not hold in practice, risking misfit with real-world data. Panayides (2010, pp. 269–288) noted that Rasch models can oversimplify educational outcomes, potentially masking differences in learner performance. Boone and Scantlebury (2006, pp. 253–269) highlighted that Rasch analysis may inadequately capture complex constructs, a concern applicable to multi-dimensional language proficiency. Tan (2024, pp. 1–17) found discrepancies between Rasch and alternative item measures, raising questions about validity, while Hope (2024, pp. 123–145) emphasized careful consideration of assumptions in diverse assessments. Overall, these studies suggest that while Rasch modeling supports scaling and comparability, its limitations must be recognized, and results interpreted cautiously for heterogeneous populations like English language learners.

Table 3

Academic Studies that have contributed to the understanding of the validity and reliability of WIDA's ACCESS assessments

	Academic Studies	Focus / relevance
1	Wolf, M. K. & Faulkner-Bond, M. (2016). Validating English Language Proficiency Assessment Uses for English Learners: Academic Language Proficiency and Content Assessment Performance. Educational Measurement: Issues and Practice, 35(2), 6-18. (ETS)	A study of how ELP test scores relate to academic language proficiency and content assessment outcomes—important validity evidence.
2	Llosa, L. (2007). Validating a Standards-Based Classroom Assessment of English Proficiency: A Multitrait-Multimethod Approach. Language Testing, 24(4), 489-515. (NYU Scholars)	Empirical work on construct validity in a classroom-based English proficiency assessment—useful methodological example.
3	Baron, P. A., Linquanti, R., & Huang, M. (2020). Validating Threshold Scores for English Language Proficiency Assessment Uses. In M. K. Wolf (Ed.), Assessing English Language Proficiency in U.S. K-12 Schools (pp. 161-184). Routledge. (ETS)	Focuses on the validity of cut-scores/thresholds in ELP assessments for K-12 English learners.
4	Faulkner-Bond, M., Wolf, M. K., Wells, C. S., & Sireci, S. G. (2018). Exploring the Factor Structure of a K–12 English Language Proficiency Assessment. Language Assessment Quarterly, 15(2), 130-149. (ETS)	A psychometric study (confirmatory factor analysis) of ELP assessment structure—reliability/validity focus.
5	Young, J. W., Cho, Y., Ling, G., Cline, F., et al. (2008). Validity and Fairness of State Standards-Based Assessments for English Language Learners. Educational Assessment, 13(2-3), 170-192. (ResearchGate)	Examines validity/fairness of large-scale, state-level standards-based assessments for ELLs—a relevant parallel.
6	Wolf, M. K., Farnsworth, T., & Herman, J. L. (2008). Validity Issues in Assessing English Language Learners' Language Proficiency. Educational Assessment, 13(2-3), 80-107. (Macrothink Institute)	A review and empirical discussion of reliability, validity and fairness issues in ELP assessments for English learners.

	Academic Studies	Focus / relevance
7	Llosa, L. (2011). Standards-based Classroom Assessments of English Proficiency: A Review of Issues, Current Developments, and Future Directions for Research. Language Testing, 28(3), 367-382. (NYU Scholars)	A review article focused on classroom-based ELP assessments, related to validity and interpretative issues.
8	Kim, M., Herman, J. L., Bachman, L. F., Bailey, A. L., & Griffin, N. (2008). Recommendations for Assessing English Language Learners: English Language Proficiency Measures and Accommodation Uses. CRESST Report 737. (UCLA) (CRESST)	This is an academic research-based guideline document on ELP assessment validity/accommodation—valuable in the literature.
9	Gutiérrez-Clellen, V. F. & Simon-Cereijido, G. (2007). The Discriminant Accuracy of a Grammatical Measure With Latino English-Speaking Children. Journal of Speech, Language & Hearing Research, 50(4), 968-981. (PMC)	While not K–12 large-scale ELP assessment, examines validity/diagnostic accuracy of a language proficiency measure in bilingual children—useful context for reliability/validity of L2 measures.
10	Cardozo-Gaibisso, L., Wood Hodges, G., Mardones-Segovia, C., & Cohen, A. S. (2024). Multidimensional Assessment Performance Analysis: A Framework to Advance Multilingual Learners' Scientific Equity in K–12 Contexts. Education Sciences, 14(10), Article 1068. (MDPI)	A recent conceptual/empirical article raising measurement, validity, reliability and equity issues in multilingual learners' assessments in K–12—relevant to ELP assessment validity debates.

These academic studies can be divided into 3 main categories:

1. Validity and Score Interpretation

ELP assessments are valid when scores reflect academic language proficiency and predict content performance. Properly validated thresholds and fair design ensure accurate classification of learners.

2. Reliability and Psychometrics

Reliable ELP assessments consistently measure language ability across items and occasions. Psychometric analyses confirm test structure and support sound instructional decisions.

3. Equity, Accommodation, and Diagnostic Utility

Assessments must be fair and accommodate diverse learners. Diagnostic measures and multidimensional frameworks help provide precise information and equitable opportunities for all students.

In brief, effective ELP assessments balance validity, reliability, and equity, providing accurate, consistent, and fair measurement for all learners. Sound psychometrics and appropriate accommodations ensure meaningful use for instruction and high-stakes decisions.

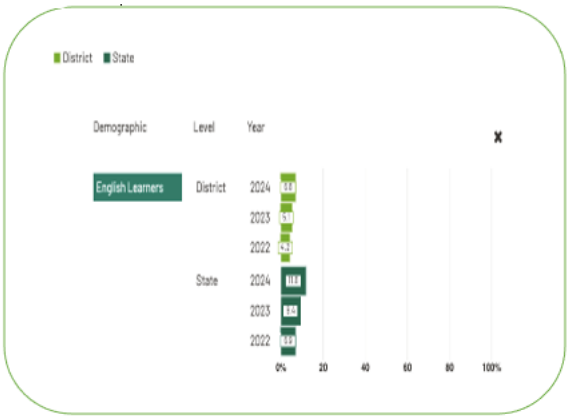
Scarcity of Conference Sessions Dedicated to Examining the Validity and Reliability of the ACCESS Test

Albeit educational and psychological assessments are widely implemented, dedicated conference sessions that examine their validity and reliability are surprisingly rare. Empirical studies reveal that many research reports neglect to present or critically assess these measurement properties. This scarcity at conferences may reflect the perception that methodological concerns are secondary, time limitations in conference programming, or the assumption that widely used instruments are inherently well-validated. Hence, opportunities to refine assessment methods, foster critical discussion, and promote best practices in instrument design are limited. While some conferences have featured sessions focused on these issues—such as the 2015 ASEE Annual Conference, which addressed the establishment of validity and reliability in observation protocols and the 2025 European Congress of Methodology, which included a session on

psychological measurement (European Association of Methodology, 2025)—these occurrences are exceptional, underscoring the need for more consistent attention to measurement rigor.

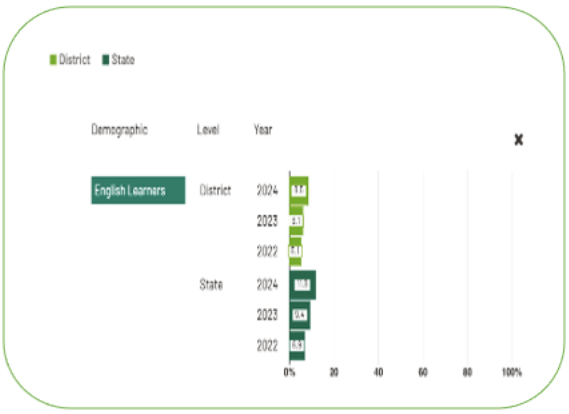
Analysis of Illinois School District Proficiency Trends: Illinois Report Card (2022–2024)

The data utilized in this study was obtained from the Illinois Report Card, a publicly accessible repository. To construct the sample, 14 school districts were randomly selected, ensuring that each district had an equal probability of inclusion. This random sampling approach was employed to minimize selection bias and enhance the representativeness of the sample within the broader population of Illinois school districts. Utilizing publicly available data also ensures transparency and allows for replication of the study by other researchers. Collectively, these methodological choices support the reliability and generalizability of the study's findings regarding educational outcomes across the state.

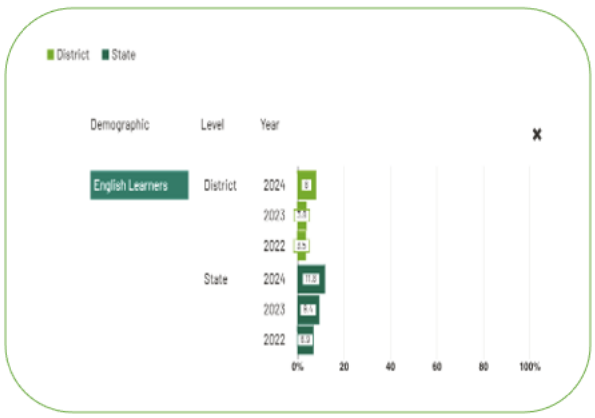


WAUKEGAN CUSD 60

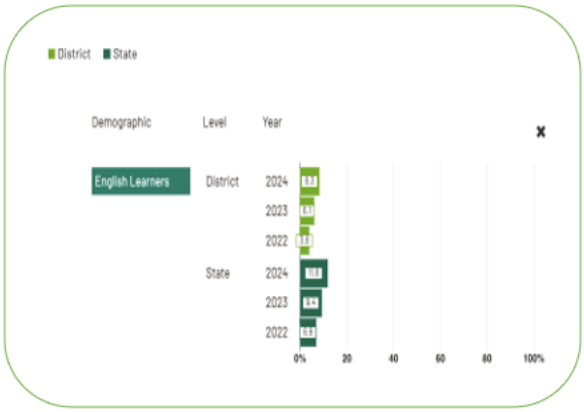
NORTH CHICAGO SD 187



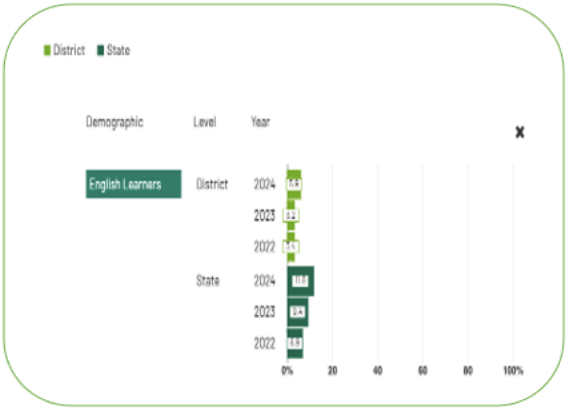
ROCKFORD SD 205



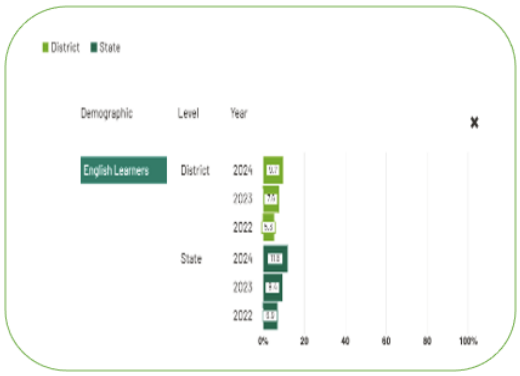
WHEELING CCSD 21



GURNEE SD 56

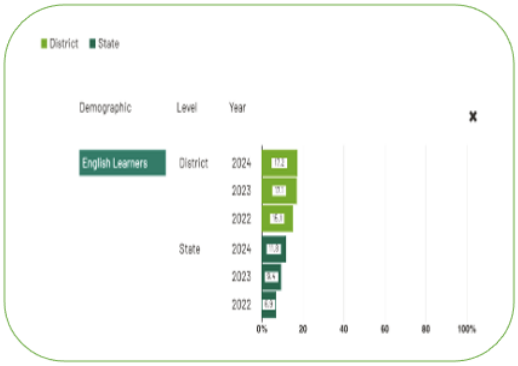


BEACH PARK CCSD 3

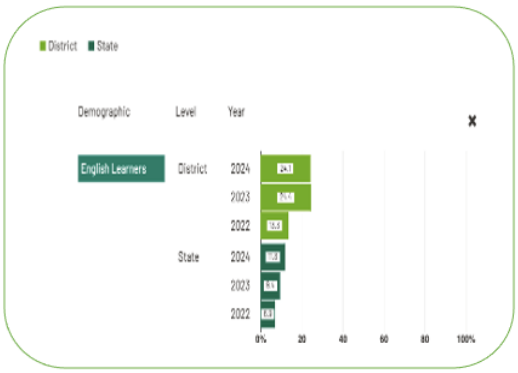


CHICAGO PUBLIC SCHOOLS DISTRICT 299

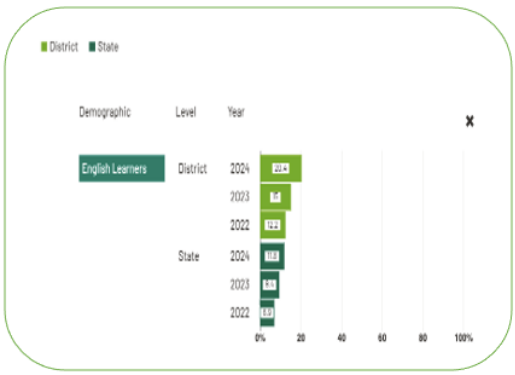
CRYSTAL LAKE CCSD 47



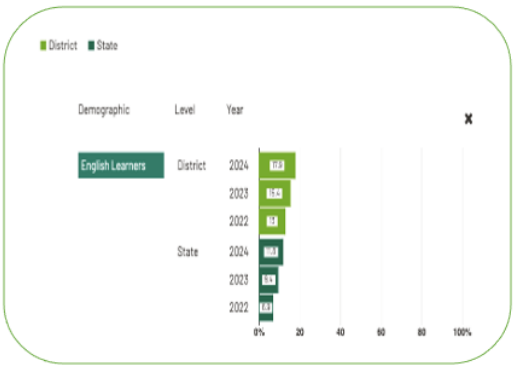
SCHAUMBURG CCSD 54



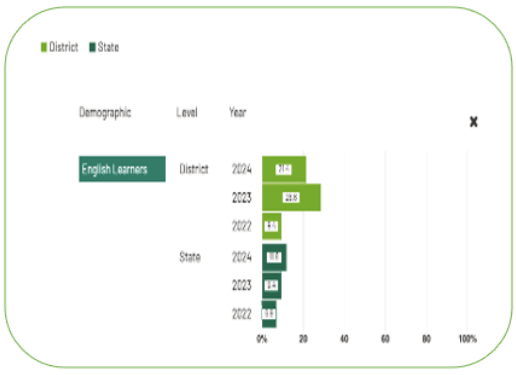
SKOKIE SD 73-5



ARLINGTON HEIGHTS SD 25



PLAINFIELD SD 202



LIBERTYVILLE SD 70

GRAYSLAKE CCSD 46

The data consistently indicate extremely slow progress in English language proficiency from 2022 to 2024. While there are minor variations, with some districts showing marginally higher gains than others (Grayslake CCSD 46, Arlington Heights 25, etc.), the overall improvement remains conspicuously minimal both at the district and state levels.

This trend suggests that, despite ongoing initiatives and targeted programs designed to support English Learners, progress toward proficiency is almost stagnant. The slow pace of development highlights systemic challenges that transcend individual district efforts, including potential limitations in the instructional core (Teaching for Biliteracy), resource allocation, or alignment with best practices in dual/bilingual education. The persistent low growth observed across various districts highlights the urgent need for a thorough evaluation of current programs, district leadership, the structure of bilingual departments, and investment in vertically aligned curricula. These factors are essential for the effective implementation and success of dual/bilingual education. Furthermore, it is pivotal to establish clear guidelines for implementing 90:10 or 50:50 bilingual programs to promote consistency and effectiveness, even though research shows that students in total immersion programs tend to achieve higher proficiency in the target language compared to those in 50:50 immersion programs (Genesee, 2004; Genesee & Lindholm-Leary, 2013).

Overall, the data suggest that, although some districts show modest gains compared to others, the prevailing trend remains minimal growth. This pattern underlines persistent barriers to effective language acquisition and calls for the urgent need for enhanced and evidence-based interventions that directly support English Learners' proficiency. In light of these findings, the researcher endeavors to problematize the data by questioning whether it truly reflects students' proficiency in classroom-based settings. Perhaps this question could serve as a starting point for further academic inquiry, guiding prospective researchers to explore the extent to which standardized data align with actual classroom proficiency.

Conclusion

The findings of this research indicate that while WIDA's ACCESS test serve as a central instrument for measuring English language proficiency in U.S. schools, its validity and reliability as a true measure of English learners' proficiency remain open to question. Evidence from prior studies and classroom-based observations reveals a persistent rupture between students' ACCESS scores and their demonstrated abilities in authentic academic and social contexts. Factors such as test design, digital administration challenges, and test-related anxiety together contribute to this gap, raising critical concerns about the fairness and instructional consequences of high-stakes testing. Moreover, the data suggest that an overreliance on standardized test scores might be conducive to narrowing instructional practices and overlooking the complex and evolving nature of bilingual and biliterate development. In essence, this study calls for a paradigm shift from perceiving ACCESS as a definitive measure of English proficiency to understanding it as one among multiple tools within a broader and context-sensitive assessment framework. In a nutshell, there should be a more inclusive and holistic representation of student language proficiency and academic potential.

Further Work

Building on these findings, future research should explore how alternative or complementary assessment models can more accurately capture EL's communicative competence across diverse contexts. A promising direction would involve theoretically developing and empirically validating a multidimensional assessment framework that integrates ACCESS test data with classroom-based performance tasks, teacher evaluations, and student self-assessments. Such a model could provide a more holistic picture of language proficiency by balancing standardized measures with authentic and context-driven indicators of second language acquisition.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219-234. https://doi.org/10.1207/s15324818ame1403_2

Al Fraidan, A. (2025). Test anxiety across writing formats: The mediating role of perceived teacher strictness. *Acta Psychologica*, *256*(104942), 104942. https://doi.org/10.1016/j.actpsy.2025.104942

Avdiu, V., & Ahmedi, V. (2024). Alternative Assessment Strategies to Enhance Learning for Students with Special Needs. *Journal of Social Studies Education Research*, *15*(5), 1–25. https://jsser.org/index.php/jsser/article/view/6024

Aydin, S., Denkci Akkaş, F., Türnük, T., Baştürk Beydilli, A., & Saydam, İ. (2020). Test anxiety among foreign language learners: A qualitative study. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2020.4686

Aydınlar, A., Mavi, A., Kütükçü, E., Kırımlı, E. E., Alış, D., Akın, A., & Altıntaş, L. (2024). Awareness and level of digital literacy among students receiving health-based education. *BMC Medical Education*, *24*(1). https://doi.org/10.1186/s12909-024-05025-w

Bachiri, H. (2025). *Teaching for Biliteracy in the United States: Pitfalls and Recommendations. ProScholar Insights, 4*(3), 78-95. https://doi.org/10.55737/psi.2025c-43106

Beeman, K. & Urow, C. (2013). *Teaching for Biliteracy: Strengthening Bridges Between Languages*. Caslon Publishing.

Bennett, R. E. (2023). Toward a theory of socio-culturally responsive assessment. *Language Testing*, 40(3), 389–411. https://doi.org/10.1080/10627197.2023.2202312

Berliner, D. C., & Nichols, S. L. (2007). High-stakes testing is putting the nation at risk. *Education Week*, *26*(27), 36-48.

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (2020). Comparability of Large-Scale Educational Assessments: Issues and Recommendations. *National Academy of Education*.

Bond TG. and Fox CM. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences.* Second ed. New York: Routledge. https://doi.org/10.4324/9781315814698

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education, 90*(2), 253–269. https://doi.org/10.1002/sce.20106

Brookhart, S. M. (2012). *SAGE handbook of research on classroom assessment.* SAGE Publications.

Celestino, M. S., Belluzzo, R. C. B., Albino, J. P., & Valente, V. C. P. N. (2024). Bibliometric analysis: Literature review and proposal of a methodological framework in 12 steps. *Revista Aracê, 6*(4), 13421–13446. https://doi.org/10.56238/arev6n4-146

Cohen, M. (2022). *8 Reasons WIDA ACCESS Student Scores May Be Invalid*. Owlcation.

Cross, R. (2025). *The impact of assessment on teaching and learning: Positive washback in classroom contexts.* Oxford University Press.

Deng, Y., & Liu, H. (2025). To overcome test anxiety in on-line assessment: unpacking the mediator roles of techno competencies, teacher support, self-efficacy, and autonomy. *BMC Psychology*, *13*(1), 192. https://doi.org/10.1186/s40359-025-02545-y

Downing, V. R. (2020). Fear of negative evaluation and student anxiety in community college students. *CBE-Life Sciences Education*, *19*(4).

Echevarria, J., Vogt, M., & Short, D. (2017). *Making content comprehensible for English The SIOP model* (5th ed.). Pearson.

Espinosa, L. (2010, March 2). *Assessment for Young ELLs: Strengths and Limitations in Current Practices*. Colorín Colorado; Colorín Colorado. https://www.colorincolorado.org/article/assessment-young-ells-strengths-and-limitations-current-practices

Foster, S. M. (2024). High-stakes Standardized Testing: Its Disproportionate Impact on Marginalized Communities.

Fox, J., & Fairbairn, S. (2011). Test review: ACCESS for ELLs®. *Language Testing, 28*(3), 425–431. https://doi.org/10.1177/0265532211404195

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*(3), 193–202. https://doi.org/10.1037/0003-066X.39.3.193

Gao, L., Zhang, Y., & Chen, J. (2025). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers & Education, 168,* 104211. https://doi.org/10.1016/j.caeai.2024.100206

Goldstein, H. (1979). Consequences of using the Rasch model for educational assessment. *British Educational Research Journal, 5*(2), 211–220. https://doi.org/10.1080/0141192790050207

Guo, S., Wang, Y., Yu, J., Wu, X., Ayik, B., Watts, F. M., Latif, E., Liu, N., & Liu, L. (2025). Artificial intelligence bias on English language learners in automatic scoring. arXiv. https://doi.org/10.48550/arXiv.2505.10643

Hope, D., Kluth, D., Homer, M., Dewar, A., Goddard-Fuller, R., Jaap, A., & Cameron, H. (2024). Exploring the use of Rasch modelling in "common content" items for multi-site and multi-year assessment. *Advances in Health Sciences Education*, *30*(2), 427-438. https://doi.org/10.1007/s10459-024-10354-y

Jalilzadeh, K., & Coombe, C. (2023). Constraints in employing learning-oriented assessment in EFL classrooms: Teachers' perceptions. *Language Testing in Asia*, *13*(1). https://doi.org/10.1186/s40468-023-00222-8

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review, 2*(2), 130–144.

Liu, H. (2025). *The mediating role of foreign language anxiety*. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC12383800/

Liu, Y., & Boone, W. J. (2006). Applying Rasch measurement in classroom assessment. *Journal of Science Education and Technology, 15*(3), 257–274.

Ly, H. H. (2024). A Review of Alternative Assessment Methods and How to Apply Them in EFL Classrooms. *Journal of Research in Humanities and Social Science, 12*(5), 176-181.

McNeal, N. L. (2016). *Correlating English language learner CRCT scores on the basis of English language learner access scores*. Liberty University.

Panayides, P. (2010). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal, 5*(2), 211–220. https://doi.org/10.1080/01411920903018182

Patterson, E., & Schneider, E. (2024). Are WIDA test results appropriately reflecting multilingual learners' language skills according to ESOL teachers' experiences?: Results of a pilot study. *GATESOL Journal*, *33*(1), 3–14. https://doi.org/10.52242/gatesol.184

Proficiency Test Series 1200–2021–2022 Administration. Madison, WI: WIDA Consortium.

Rasch, G. (1960). *Probabilistic model for some intelligence and achievement tests.* Copenhagen: Danish Institute for Educational Research.

Salazar, M. C. (2022). An Analysis of Variables Impacting English Learner Achievement on Science Assessments. *Sustainability, 14*(13), 7814. https://www.mdpi.com/2071-1050/14/13/7814

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests.* Longman

Smith, T. E. (2022). Self-management interventions for reducing challenging classroom behaviors. *Campbell Systematic Reviews, 18*(1), e1234. https://doi.org/10.1002/cl2.1223

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher (Washington, D.C.: 1972)*, *32*(2), 3–13. https://doi.org/10.3102/0013189x032002003

State Superintendent of Education. (2020). *2019-2020 Accessibility and Accommodations Supplement.*

Tan, L. Y., McLean, S., Kim, Y. A., & Vitta, J. P. (2024). Rasch modelling vs. item facility: implications on the validity of assessments of Asian EFL/ESL vocabulary knowledge and lexical sophistication modelling. *Language Testing in Asia*, *14*(1). https://doi.org/10.1186/s40468-024-00327-8

Villegas, L. (2022, October 19). *Long-standing Limitations of English Learner Academic Assessment Data Persist*. New America. https://www.newamerica.org/education-policy/edcentral/long-standing-limitations-of-english-learner-academic-assessment-data-persist/

Waters, C. N. (2020). *Teachers' perceptions of the broad validity of a high stakes English language proficiency test* [Dissertation, Virginia Commonwealth University]. Scholars Compass.

Waters, C. N. (2021). *Considerations for Using ACCESS Test Scores in Decision Making.* Virginia Commonwealth University Scholars Compass.

WIDA Consortium. (2011). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 100–2004–2005 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2013). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 400–2012–2013 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2014). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 500–2014–2015 Administration*. Madison, WI: WIDA Consortium.

WIDA Consortium. (2015). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 600–2015–2016 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2016). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 700–2016–2017 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2017). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 800–2017–2018 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2018). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 900–2018–2019 Administration*. Madison, WI: WIDA Consortium.

WIDA Consortium. (2019). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1000–2019–2020 Administration*. Madison, WI: WIDA Consortium.

WIDA Consortium. (2020). *ACCESS for ELLs 2.0: Test administration manual.* WIDA.

WIDA Consortium. (2020). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1100–2020–2021 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2021). *Annual Technical Report for ACCESS for ELLs® English Language*

WIDA Consortium. (2022). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1300–2022–2023 Administration*. Madison, WI: WIDA Consortium.

WIDA Consortium. (2023). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1400–2023–2024 Administration.* Madison, WI: WIDA Consortium.

WIDA Consortium. (2024). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test Series 1500–2024–2025 Administration*. Madison, WI: WIDA Consortium.

WIDA Consortium. (2025). *Annual Technical Report for WIDA Alternate ACCESS for ELLs.* Madison, WI: WIDA Consortium.

WIDA. (2019). *Investigating K–12 English learners' use of universal tools embedded in online language assessments.*

WIDA. (2020). *WIDA Accessibility and Accommodations Framework.* https://wida.wisc.edu/assess/accessibility-accommodations

WIDA. (2025). *ACCESS for ELLs Interpretive Guide for Score Reports Grades K-12.* Board of Regents of the University of Wisconsin System.

WIDA. 2019-2020 Accessibility and Accommodations Supplement.

Wolf, M. K. (2008). Validity Issues in Assessing English Language Learners' Language Proficiency. *Language Testing, 25*(3), 355–377. https://doi.org/10.1080/10627190802394222

Wolf, M. K. (2010). Improving the Validity of English Language Learner Assessments. *Educational Policy, 24*(4), 523–545. https://www.tandfonline.com

Zheng, Y., & Cheng, L. (2018). How does anxiety influence language performance? From the perspectives of foreign language classroom anxiety and cognitive test anxiety. *Language Testing in Asia*, *8*(1). https://doi.org/10.1186/s40468-018-0065-4