

Reconceptualizing Assessment in Science of Reading—Aligned TESOL Contexts: Validity, Equity, and Linguistic Responsiveness in the Middle East

Salam Sammour¹, Mohammed Al-Badawi², Aladdin Al-Tarawneh³, Ala Yaghi⁴

¹ English Language Center – University of Sharjah, Sharjah, United Arab Emirates

² Department of English Language and Literature—AlZaytoonah University of Jordan, Amman, Jordan

³ Department of English Language and Literature – Zarqa University, Zarqa, Jordan

⁴ College of Arts, Humanities, and Social Sciences – University of Kalba, Sharjah, United Arab Emirates

Correspondence: Mohammed Al-Badawi, Department of English Language and Literature—AlZaytoonah University of Jordan, Amman, Jordan.

Received: February 6, 2026

Accepted: May 14, 2026

Online Published: October 13, 2026

doi:10.5430/wjel.v16n6p631

URL: <https://doi.org/10.5430/wjel.v16n6p631>

Abstract

The global expansion of the Science of Reading (SoR) has reshaped literacy instruction through a renewed emphasis on systematic, evidence-based approaches to foundational reading skills. While SoR-informed instructional reforms have been widely implemented across diverse educational systems, assessment practices—particularly in English as a Foreign Language (EFL) context—have received comparatively limited critical scrutiny. In the Middle East, English literacy development occurs within linguistically complex environments characterized by Arabic diglossia, uneven exposure to English, and stratified schooling systems. Despite these realities, SoR-aligned reading assessments are frequently transferred from first-language English contexts with minimal adaptation, often carrying implicit assumptions about language exposure, phonological development, and reading trajectories.

This article adopts a conceptual and critical review approach grounded in theoretical synthesis. Drawing on interdisciplinary scholarship from the Science of Reading, bilingual literacy development, second language acquisition, and assessment validity theory, the article critically examines the applicability of SoR-aligned assessment practices in multilingual TESOL contexts and proposes a linguistically responsive framework for equitable assessment interpretation and use.

The article argues that unadopted assessment practices risk conflating normal second-language development with reading difficulty. Such misinterpretations may lead to over-identification, inappropriate intervention, and inequitable educational outcomes, particularly for learners with limited access to English outside formal schooling. The article proposes a linguistically responsive orientation to SoR-aligned assessment that preserves diagnostic rigor while accounting for cross-linguistic influence, exposure variability, and contextual constraints. By foregrounding assessment interpretation, teacher professional judgment, and policy implications, the article contributes to ongoing debates about the ethical and effective implementation of SoR-informed literacy reforms in multilingual TESOL settings.

Keywords: Science of Reading, assessment validity, TESOL, Arabic-speaking EFL learners, multilingual literacy

1. Introduction

Over the past decade, the Science of Reading (SoR) has emerged as a dominant framework shaping literacy policy and practice across multiple educational systems. Rooted in decades of interdisciplinary research in cognitive psychology, linguistics, neuroscience, and education, SoR emphasizes the systematic development of foundational reading skills—particularly phonological awareness, decoding, orthographic mapping, and fluency—as necessary conditions for skilled reading comprehension (Castles et al., 2018; Ehri, 2020). In first-language English contexts, these principles have been widely promoted as a corrective to instructional approaches perceived as insufficiently explicit or theoretically grounded.

As SoR-informed approaches have gained international traction, they have increasingly been adopted beyond their original Anglophone contexts. In the Middle East, ministries of education, accreditation agencies, and private education providers have embraced structured literacy programs aligned with SoR principles, often framing them as evidence-based solutions to persistent concerns about English proficiency and performance on international assessments. English literacy, in this context, is frequently positioned as a gateway to higher education access, labor market participation, and global competitiveness (Snow, 2002).

However, the global diffusion of instructional frameworks raises critical questions about contextual alignment—particularly with respect to assessment. Within the SoR paradigm, assessment is not ancillary; it plays a central role in identifying learners at risk,

guiding instructional decisions, and determining eligibility for intervention (Kilpatrick, 2015). In many educational systems, assessment outcomes also carry high-stakes consequences, shaping student placement, academic streaming, and long-term educational trajectories. Despite this centrality, assessment practices have often been transferred with considerably less scrutiny than instructional methods, especially in TESOL contexts.

In Middle Eastern EFL settings, English literacy development unfolds under sociolinguistic conditions that differ substantially from those assumed by most SoR-aligned assessment tools. Learners typically acquire literacy in Arabic—a language characterized by diglossia and a relatively transparent consonantal orthography—before encountering English as an additional language (Saiegh-Haddad & Joshi, 2014; Share, 2008). Exposure to English varies widely depending on school type, socioeconomic background, and home language practices. Learners enrolled in international or private schools may receive extensive English input from early childhood, while those in public or under-resourced schools may encounter English primarily through formal instruction with limited opportunities for meaningful language use (Cummins, 2000; Snow, 2002).

These disparities have direct implications for reading development and assessment interpretation. Performance on English reading tasks reflects not only cognitive and linguistic skill, but also the quantity and quality of exposure to English. Features such as reduced reading fluency, difficulty with vowel contrasts, or reliance on first-language phonotactic patterns may reflect typical second-language development rather than underlying reading difficulty (Perfetti, 2017; Share, 2008). When assessment tools fail to account for these realities, there is a heightened risk that assessment outcomes will misrepresent learners' abilities.

Such misinterpretations are not merely technical concerns. Assessment results often inform instructional decisions that shape learners' educational experiences and identities. Over-identification of EFL learners as struggling readers can lead to unnecessary remediation, lowered expectations, and restricted access to grade-level content. Conversely, failure to distinguish between linguistic difference and genuine reading difficulty undermines both educational equity and the ethical use of assessment data (Messick, 1995; Snow, 2002).

This article argues that while the instructional principles of the Science of Reading offer valuable insights for TESOL contexts, assessment practices must be reconceptualized to ensure validity, equity, and ethical responsibility. Focusing on Arabic-speaking EFL learners in the Middle East, the article critically examines the assumptions embedded in commonly used SoR-aligned assessments, explores their limitations in multilingual contexts, and proposes a linguistically responsive framework for assessment interpretation and use. By centering assessment—an often under-theorized component of SoR implementation—the article contributes to broader discussions in reading science, language assessment, and TESOL regarding the responsible application of evidence-based literacy reforms in linguistically diverse settings.

The aim of this study is to critically examine the applicability of Science of Reading-aligned assessment practices within multilingual TESOL contexts, particularly among Arabic-speaking EFL learners in the Middle East. The article further seeks to explore how monolingual assumptions embedded within SoR-aligned assessment tools may influence validity, equity, and interpretation when applied in linguistically diverse educational environments. Accordingly, to guide this conceptual analysis, the article seeks to address the following research questions:

1. How do Science of Reading-aligned assessment practices align with the linguistic realities of Arabic-speaking EFL learners?
2. What challenges emerge when monolingual English reading assessment frameworks are transferred to multilingual TESOL contexts without sufficient adaptation?
3. How can linguistically responsive assessment practices improve validity, equity, and ethical interpretation within SoR-informed literacy reform?

The theoretical foundation of this article is situated at the intersection of reading science, second language acquisition, and assessment validity. Moreover, the Science of Reading does not represent a single, unified theory of literacy development. Rather, it constitutes a synthesis of converging evidence drawn from cognitive psychology, linguistics, neuroscience, and educational research. Central to this synthesis is the assertion that skilled reading requires the coordinated development of multiple component skills and that difficulties in any of these components can impede comprehension. While these claims are well supported in first-language contexts, their application to multilingual and EFL settings requires careful theoretical integration.

For instance, the Simple View of Reading (SVR) in Second-Language contexts is one of the most influential models underpinning the Science of Reading which conceptualizes reading comprehension as the product of decoding and linguistic comprehension (Tunmer & Hoover, 2019). According to this model, weaknesses in either component can constrain reading outcomes, and assessment practices are designed to identify the source of difficulty so that instruction can be targeted accordingly.

In monolingual English-speaking contexts, the SVR has proven robust in explaining patterns of reading development and difficulty. However, when applied to EFL learners, the model behaves differently in practice. While decoding skills may develop at a measurable pace through explicit instruction, linguistic comprehension in the target language is often constrained by limited vocabulary, reduced oral language exposure, and uneven opportunities for language use. As a result, EFL learners may demonstrate adequate decoding skills while continuing to struggle with comprehension, or conversely, show weak decoding performance that reflects limited exposure rather than underlying difficulty.

This asymmetry poses challenges for assessment interpretation. When SVR-aligned assessments identify weaknesses in decoding among EFL learners, it is not always clear whether these weaknesses reflect inefficient orthographic mapping or simply limited familiarity with the phonological and orthographic patterns of English. Without attention to learners' language histories, assessment outcomes may be interpreted through a monolingual lens that overstates deficit and understates developmental variation.

Second language acquisition research underscores this concern by emphasizing that language proficiency develops gradually and unevenly across domains. Vocabulary growth, morphosyntactic development, and phonological sensitivity may progress at different rates depending on exposure conditions and instructional context (Cummins, 2000). In EFL settings, where input is often restricted to classroom interaction, these developmental trajectories may extend well beyond the timelines assumed in monolingual reading models.

Furthermore, Orthographic mapping —the process through which readers form automatic connections between phonemes, graphemes, and meaning—is another core construct within the Science of Reading (Ehri, 2020). Efficient orthographic mapping enables fluent word recognition and reduces cognitive load during reading. In monolingual contexts, this process is supported by extensive exposure to spoken and written language, allowing learners to consolidate phonological representations and build a large sight vocabulary.

For EFL learners, orthographic mapping unfolds under markedly different conditions. Exposure to English print may be limited, intermittent, or heavily mediated by instructional materials. Phonological representations of English words may be incomplete or influenced by first-language phonotactic patterns. These factors can slow the development of automaticity without indicating underlying cognitive impairment.

Cross-linguistic research highlights that learners approach new writing systems through the lens of their first language. Arabic-speaking learners, for example, come to English with experience in a consonantal writing system in which short vowels are often omitted and lexical distinctions rely less heavily on vowel representation. English, by contrast, depends heavily on vowel contrasts to signal meaning and grammatical distinctions. As a result, predictable error patterns—such as vowel substitution, reduction, or omission—are common among Arabic-speaking EFL learners (Share, 2008; Perfetti, 2017).

These patterns are developmentally expected and reflect transfer rather than deficit. However, SoR-aligned assessments that treat such errors as indicators of phonological weakness risk misinterpreting normal bilingual development. Without explicit consideration of cross-linguistic influence, assessment tools may overestimate the prevalence of decoding difficulties among EFL learners and underestimate their capacity to improve with increased exposure and instruction.

Second language acquisition theory and literacy development further complicates the application of SoR principles to EFL assessment. Cummins' Interdependence Hypothesis proposes that literacy-related skills developed in a first language can transfer to a second language, particularly when learners have achieved a threshold level of proficiency (Cummins, 2000). This suggests that strong first-language literacy can support second-language reading development, even when surface-level performance appears weak.

However, transfer is neither automatic nor uniform. Structural differences between languages mediate how literacy skills are applied, and the extent of transfer depends on instructional context and exposure. In Arabic–English contexts, differences in orthographic depth, vowel representation, and morphological structure shape how learners approach decoding and word recognition. These differences must be accounted for when interpreting assessment data.

Research on bilingual literacy development consistently emphasizes that reading behaviors must be interpreted in light of learners' linguistic repertoires. Errors that appear problematic from a monolingual perspective may represent strategic adaptations to unfamiliar orthographic patterns. Assessment practices that ignore this complexity risk conflating language learning with reading disorder.

In order to discuss the assessment validity in multilingual contexts, it is observed that contemporary theories of assessment validity provide a crucial framework for integrating insights from reading science and second language acquisition. Validity is no longer conceptualized as a static property of a test, but as the degree to which evidence and theory support the interpretations and uses of assessment results (Messick, 1995). This shift places interpretive responsibility at the center of assessment practice.

In multilingual contexts, construct validity is particularly vulnerable to distortion. Tasks designed to measure decoding efficiency or phonological awareness may inadvertently tap language exposure, vocabulary knowledge, or familiarity with specific phonemes. When such construct-irrelevant variance influences scores, assessment outcomes become difficult to interpret meaningfully.

Consequential validity further underscores the ethical dimension of assessment use. Assessment results shape instructional placement, access to intervention, and learners' academic identities. When EFL learners are systematically misclassified due to inappropriate assessment interpretation, the consequences extend beyond individual classrooms to broader patterns of educational inequality (Snow, 2002).

Integrating reading science with second language acquisition and validity theory therefore requires moving beyond technical accuracy toward contextualized interpretation. Assessment practices must be evaluated not only in terms of what they measure, but also in terms of how results are used and whose interests they serve.

2. Assessment Assumptions and Sources of Misinterpretations

Within the Science of Reading framework, assessment is designed to function as a precise diagnostic mechanism. Measures of phonemic awareness, decoding accuracy, reading fluency, and spelling are intended to isolate component skills and identify the source of reading difficulty so that instruction can be targeted efficiently (Kilpatrick, 2015). In monolingual English-speaking contexts, these assessments have demonstrated strong predictive validity and have contributed to earlier identification of reading difficulties.

However, the effectiveness of SoR-aligned assessment tools depends on a set of assumptions about learners' linguistic experiences. These tools presuppose extensive exposure to English phonology, familiarity with phoneme inventories, and consistent opportunities for oral and written language interaction. In TESOL contexts, particularly in the Middle East, these assumptions are frequently violated.

EFL learners often encounter English primarily through formal instruction, with limited opportunities for spontaneous language use beyond the classroom. As a result, performance on phoneme-level tasks may reflect unfamiliarity with specific English sounds rather than inefficiency in phonological processing.

When assessment scores are interpreted without reference to exposure history, low performance is easily—and incorrectly—attributed to underlying reading difficulty.

Benchmark-driven assessment practices further exacerbate this problem. Many SoR-aligned assessment systems rely on normative cut scores derived from monolingual populations. These benchmarks are often treated as universal indicators of risk, despite the fact that they reflect developmental trajectories shaped by extensive English exposure. Applying such benchmarks to EFL learners risks systematic misclassification, particularly in early stages of English literacy development.

The reliance on static assessment formats compounds these challenges. Single-time-point measures provide limited insight into learners' capacity for growth or responsiveness to instruction. For EFL learners, whose performance may change rapidly with increased exposure, static assessments can underestimate learning potential and overestimate deficit. Without complementary approaches that capture developmental change, assessment outcomes remain vulnerable to misinterpretation (Ehri, 2020).

3. Over-Identification, False Positives, and the Consequences of Assessment Misuse

Building on the assessment assumptions already discussed, one consequence of misinterpretation in multilingual contexts is over-identification. Learners may be falsely classified as having reading difficulties if assessment outcomes are interpreted without considering language exposure, cross-linguistic influence, and developmental differences. These false positives are not only a measurement problem but an educational concern.

The problem of over-identification is intensified in systems that implement tiered intervention frameworks such as Response to Intervention (RTI) or Multi-Tiered Systems of Support (MTSS). These frameworks assume that inadequate response to instruction signals underlying difficulty. In EFL contexts, however, limited response may reflect insufficient language input rather than instructional mismatch or cognitive impairment (Kilpatrick, 2015).

4. Linguistic Contexts of Arabic-Speaking EFL Learners

Arabic-speaking EFL learners acquire literacy within sociolinguistic environments that differ markedly from those assumed by most English reading assessments. One of the most defining features of this context is diglossia—the coexistence of colloquial spoken dialects and Modern Standard Arabic (MSA), the formal variety used in literacy, education, and official domains. For many learners, formal schooling represents their first sustained exposure to the variety of Arabic used in print.

This diglossic situation places additional cognitive and linguistic demands on learners. Children must learn to map spoken language to print in a linguistic variety that differs from their home dialect, often while simultaneously developing literacy in a second language. These layered demands shape phonological awareness, lexical access, and metalinguistic development prior to the introduction of English (Share, 2008).

When English is introduced as an additional language, learners must navigate yet another orthographic system—one that differs substantially from Arabic in depth, vowel representation, and morphological structure. English relies heavily on vowel contrasts to signal meaning, whereas Arabic orthography places less emphasis on short vowels. As a result, Arabic-speaking learners frequently demonstrate predictable patterns when reading English, including vowel substitution, reduction, or omission (Perfetti, 2017).

These patterns are developmentally expected and reflect cross-linguistic transfer rather than disorder. However, SoR-aligned assessments that treat such errors as indicators of phonological weakness risk misclassifying learners. Without explicit recognition of transfer effects, assessment tools may overestimate the prevalence of decoding difficulties among Arabic-speaking EFL learners.

Exposure variability further complicates assessment interpretation. Learners in international or private schools often receive extensive English input across academic subjects and social contexts, while those in public or under-resourced schools may encounter English primarily through limited instructional hours. These disparities shape reading development in ways that are not captured by standardized assessments.

Research with Arabic-speaking EFL learners demonstrated that differences in English reading performance are due to differences in

language exposure, instructional quality, and opportunities for language use. Studies conducted in the Arab Gulf have shown a difference in reading fluency and comprehension outcomes based on how much they interact with the language. Al-Seghayer (2019) says that Saudi EFL learners' first source of difficulty with lower-level processing is their insufficient vocabulary knowledge and inability to use word-learning strategies. This leads to EFL learners struggling with word recognition and reading fluency. "Because their cognitive effort is focused on decoding rather than on understanding the text, these learners are compromised in their ability to completely comprehend the text." (Al-Seghayer, 2019, p.500)

Al-Qahtani (2016) also investigated the reasons behind Saudi learners' poor reading skills and found that most Saudi EFL students lack the necessary reading habits in L1 and L2, and that they lack exposure to the target language, as well as poor teaching skills and teacher training programs, and limited opportunities for reading skills development.

Failure to account for diglossia, orthographic contrast, and exposure variability increases the likelihood that assessment practices will misrepresent learners' abilities. Reading behaviors that reflect normal second-language development may be interpreted as evidence of deficit, triggering inappropriate instructional responses.

5. Toward Linguistically Responsive Science of Reading–Aligned Assessment

Rather than rejecting the Science of Reading or its associated assessment practices, this article proposes some recommendations and conceptual frameworks which argues for a reorientation toward linguistic responsiveness. A linguistically responsive approach seeks to preserve the diagnostic strengths of SoR-aligned assessment while aligning interpretation and use with the realities of second-language development. Such an approach recognizes that reading development is shaped not only by cognitive processes but also by language exposure, cross-linguistic influence, and educational context.

The first pillar of linguistically responsive assessment involves Construct Alignment. According to this, assessment constructs must be grounded in an understanding of how reading-related skills manifest in multilingual learners. In Arabic-speaking EFL contexts, this requires explicit recognition of cross-linguistic contrasts between Arabic and English phonology, orthography, and morphology.

For example, tasks assessing vowel discrimination or manipulation must be interpreted with awareness of Arabic's reduced emphasis on short vowels in print. Errors related to vowel representation should be understood as developmentally expected transfer effects rather than indicators of phonological deficit (Share, 2008; Perfetti, 2017).

Arabic orthography usually omits short vowels in print, necessitating readers to rely on morphological and contextual information for word recognition (Abu-Rabia 2019; Bar-on et al. (2018). Phonological substitutions seen in EFL learners may reflect interference from L1 reading systems and not reading impairment (Koda, 2005) Also, research on L2 reading development suggests that learners learn on a different curve than monolingual English readers because of exposure and processing demands (Melby-Lervåg et al. 2014). Before labeling an L2 with reading difficulties, their exposure history and progress over time should be taken into consideration.

A second pillar of linguistic responsiveness is the incorporation of dynamic assessment approaches alongside static measures. Dynamic Assessment emphasizes learners' responsiveness to instruction, providing insight into learning potential rather than fixed performance. This approach is particularly valuable in EFL contexts, where initial performance may underestimate capacity due to limited exposure to the target language (Lantolf & Poehner, 2014).

A second pillar of linguistic responsiveness is the incorporation of dynamic assessment approaches alongside static measures. Dynamic Assessment emphasizes learners' responsiveness to instruction, providing insight into learning potential rather than fixed performance. This approach is particularly valuable in EFL contexts, where initial performance may underestimate capacity due to limited exposure to the target language (Lantolf & Poehner, 2014).

Within a SoR-aligned framework, dynamic assessment can complement traditional measures of decoding and phonological awareness by examining how learners respond to targeted support. Learners who demonstrate rapid improvement when provided with explicit instruction may require increased exposure rather than intensive remediation. Conversely, learners who struggle to benefit from instruction despite adequate exposure may warrant further evaluation. By shifting the focus from static scores to developmental trajectories, dynamic assessment reduces the risk of false positives and supports more equitable identification practices.

Rather than relying on a single decoding score, educators may use different approaches where instruction and assessments are integrated. According to Lantof & Poehner (2014) and Poehner (2008) a student's potential for learning can be discovered through mediated assistance. This allows educators to observe performance changes after explicit instruction. Fuchs & Fuchs (2006) show that learners who demonstrate improvement after targeted support do not have reading impairments. They are probably experiencing instructional or exposure-related difficulties. Koda (2005) also found that errors in L2 reading may reflect transfer from L1 rather than reflect a deficit in decoding ability.

The third pillar of linguistically responsive assessment is Teacher Professional Judgment and Assessment Literacy which highlights the central role of teacher professional judgment. Assessment tools do not interpret themselves; their educational impact depends on how results are understood and acted upon. In multilingual contexts, teachers' knowledge of reading development, second language acquisition, and cross-linguistic influence is essential for valid interpretation.

Teacher assessment literacy involves more than technical familiarity with instruments. It includes the ability to integrate multiple

sources of evidence, recognize the limitations of standardized measures, and situate assessment results within learners' linguistic and instructional histories (Snow, 2002). When teachers are positioned solely as implementers of assessment protocols, opportunities for nuanced interpretation are diminished. Linguistically responsive systems therefore require investment in professional development that equips teachers to exercise informed judgment.

Professional judgment should not be conflated with subjectivity. Rather, it represents theory-informed decision-making that enhances validity by contextualizing assessment data. In EFL contexts, teacher judgment functions as a critical safeguard against mechanical interpretation and deficit-oriented classification.

Teachers working in multilingual classrooms may consider reviewing students' English exposure history before interpreting their assessments. Cummins (2000) suggests that language performance is strongly influenced by opportunities for exposure and instruction. They can also consider cross-linguistic transfer patterns when evaluating reading errors and monitor their growth over time before making referral decisions. Koda (2005) finds that this transfer may shape learners' reading errors. These practices support a more valid interpretation of assessment data.

6. Implications for Policy and Teacher Education

The adoption of SoR-aligned literacy reforms in the Middle East has significant policy implications, particularly with respect to assessment. Policymakers must recognize that instructional frameworks and assessment systems do not travel equally well across contexts. While structured literacy instruction may be adapted with relative success, assessment practices require far greater contextualization.

Policy frameworks that mandate SoR-aligned assessment without explicit guidance on interpretation risk reinforcing inequities. In systems where English proficiency functions as a gatekeeper to academic and economic opportunity, misclassification based on inappropriate assessment use can have long-term consequences for learners. Policymakers must therefore balance accountability with flexibility, ensuring that assessment systems support rather than constrain equitable literacy development.

Teacher education represents a critical leverage point in this process. Preparing teachers to implement structured literacy instruction without equivalent training in assessment interpretation creates an imbalance that undermines reform efforts. Pre-service and in-service programs must integrate reading science, second language acquisition, and assessment literacy into a coherent preparation framework. Case-based learning, analysis of multilingual assessment data, and opportunities for collaborative interpretation can support the development of professional judgment.

To support linguistically responsive assessment practices, teacher education programs should include training in L2 acquisition (Lucas et al., 2008), analysis of authentic multilingual learner assessment data, case studies showing assessment interpretation challenges, and opportunities for collaborative discussion of assessment decisions. Such preparation would strengthen teachers' ability to interpret assessment results.

Accreditation bodies and inspection frameworks also play a powerful role in shaping assessment practices. When accountability systems prioritize numerical benchmarks without regard for linguistic diversity, schools may feel compelled to use assessments in ways that are misaligned with learners' needs. Greater alignment between accountability expectations and linguistic realities is essential for sustainable reform.

7. Assessment, Power, and Educational Equity in Multilingual Education Systems

Assessment practices do not operate in neutral educational spaces. Rather, they are embedded within broader systems of power, accountability, and policy that shape whose knowledge is valued and whose abilities are recognized. In multilingual education systems, particularly those in postcolonial or globally stratified contexts, assessment often functions as a mechanism through which linguistic hierarchies are reproduced. The use of SoR-aligned assessments in TESOL contexts must therefore be examined not only for technical validity, but also for their role in structuring access, opportunity, and legitimacy.

In the Middle East, English occupies a powerful symbolic and material position. Proficiency in English is frequently associated with academic success, social mobility, and global participation. As a result, English literacy assessments carry disproportionate weight in shaping students' educational trajectories. When SoR-aligned assessments are applied without linguistic adaptation, they may inadvertently reinforce existing inequalities by privileging learners who have greater access to English-rich environments (Snow, 2002).

Research on educational assessment has long emphasized that tests reflect the values and assumptions of the systems in which they are developed (Messick, 1995). In the case of SoR-aligned assessments, these assumptions include monolingual norms of language exposure, linear developmental trajectories, and early mastery of phoneme-grapheme correspondences. When such assumptions are imposed on multilingual learners, assessment outcomes may reflect social advantage rather than reading ability.

This dynamic is particularly evident in stratified schooling systems. Learners in international or private schools, who often receive extensive English exposure across subjects, tend to perform well on SoR-aligned assessments. In contrast, learners in public or under-resourced schools—whose exposure to English may be limited to a few instructional hours per week—are more likely to be identified as at risk. Without contextualized interpretation, assessment results risk functioning as proxies for socioeconomic status

rather than indicators of difficulty of reading.

The consequences of such practices extend beyond individual learners. At the institutional level, assessment data are often used to evaluate school effectiveness, inform policy decisions, and justify resource allocation. When assessment systems systematically disadvantage certain learner populations, they contribute to cycles of under-resourcing and deficit labeling. In this way, assessment becomes a tool of stratification rather than support.

Equity-oriented assessment requires a shift from uniform benchmarks toward context-sensitive interpretation. This does not imply abandoning standards or lowering expectations. Rather, it involves recognizing that equitable assessment must account for differential opportunities to learn. From this perspective, fairness is achieved not through identical measurement, but through interpretation that acknowledges linguistic and educational realities (Snow, 2002).

8. A Conceptual Model for Linguistically Responsive SoR-Aligned Assessment in TESOL Contexts

To operationalize the principles discussed throughout this article, this section proposes a conceptual model for linguistically responsive SoR-aligned assessment in TESOL contexts. The model integrates insights from reading science, second language acquisition, and assessment validity theory to guide interpretation and decision-making rather than test design alone. The core components of the proposed model rests on four interrelated components: exposure awareness, cross-linguistic interpretation, developmental sensitivity, and professional judgment.

First, exposure awareness requires that assessment interpretation explicitly account for learners' opportunities to engage with English in oral and written forms. Assessment data should never be interpreted independently of information about instructional time, language of schooling, and access to English beyond the classroom. Research consistently demonstrates that exposure is a primary driver of second-language literacy development (Cummins, 2000; Snow, 2002).

Second, cross-linguistic interpretation involves recognizing predictable transfer patterns between Arabic and English. Errors related to vowel representation, consonant clusters, or morphological endings should be evaluated in light of known contrasts between the two languages rather than treated as evidence of phonological deficit (Share, 2008; Perfetti, 2017). This component directly addresses the risk of construct-irrelevant variance in assessment interpretation.

Third, developmental sensitivity emphasizes trajectories over thresholds. Rather than relying solely on static benchmarks, educators should examine patterns of growth across time. Learners who demonstrate steady progress with increased exposure and instruction may not require intensive intervention, even if they initially perform below monolingual norms. This aligns with research emphasizing growth-oriented assessment practices in multilingual contexts (Lantolf & Poehner, 2014).

Finally, professional judgment serves as the integrating mechanism of the model. Teachers synthesize assessment data, classroom evidence, and linguistic knowledge to make informed decisions. This form of judgment is not subjective intuition but theory-informed interpretation grounded in empirical research (Snow, 2002). This paper proposes implications for identification and intervention; with regards to identification practices, the model cautions against premature labeling of EFL learners as struggling readers. Identification decisions should be made only after learners have had sufficient opportunity to develop English literacy skills and after alternative explanations for performance—such as limited exposure—have been carefully considered. This approach reduces the likelihood of false positives and supports more targeted use of intervention resources.

Evidence from multilingual assessment research supports this argument. Studies have shown that reading difficulties that are discovered with initial screening may be resolved after language exposure and targeted instructional support. Vaughn et al. (2019) claim that reading difficulties in English learners may improve after exposure to language and targeted instructional interventions. Her results indicated that students who received the reading intervention showed greater improvement in reading outcomes compared to the control group. Similar findings were expressed in 2017 by Ritchey et al, who found that students who received the intervention showed improved reading comprehension outcomes. The intervention came in the form of targeted instruction that included teaching comprehension strategies, understanding main ideas and details, working with informational text structure, and guided practice.

Such findings reinforce the importance of thinking about responsiveness to instruction and language learning opportunities before making identification decisions. As such, adaptive approaches are important to help distinguish temporary language-related challenges from persistent reading-related difficulties.

For intervention design, the model emphasizes alignment between instructional focus and underlying need. Learners whose difficulties stem primarily from limited exposure may benefit more from increased language-rich instruction than from narrowly targeted decoding intervention. Conversely, learners who fail to respond to instruction despite adequate exposure may warrant further evaluation for reading-related difficulties.

At the system level, the model provides a framework for aligning assessment policy with linguistic reality. Ministries of education and accreditation bodies can use the model to evaluate whether existing assessment practices support or undermine equity goals. Professional development initiatives can be designed to build teachers' capacity to apply the model consistently across contexts.

By shifting the focus from test scores to interpretation practices, the model reframes assessment as a tool for support rather than classification. This reframing is essential for ensuring that SoR-aligned reforms fulfill their promise in multilingual TESOL settings.

While this article has advanced a conceptual and theoretical critique of SoR-aligned assessment practices in TESOL contexts, it is essential to address the methodological implications of this work. One of the persistent challenges in debates surrounding the Science of Reading is the tendency to privilege certain forms of evidence, particularly large-scale quantitative studies—while marginalizing forms of inquiry that capture contextual and developmental complexity. In multilingual EFL contexts, this imbalance has direct consequences for how assessment practices are validated, interpreted, and justified.

Much of the empirical research underpinning SoR-aligned assessment tools is derived from monolingual English-speaking populations. These studies have generated valuable insights into decoding development, phonological processing, and early reading trajectories. However, their methodological assumptions—particularly regarding language exposure, instructional consistency, and developmental timelines—are not always transferable to EFL settings. When such evidence is applied uncritically, it can obscure the role of contextual variables that are central to multilingual literacy development.

Research on TESOL assessment must therefore adopt methodological designs that explicitly account for linguistic diversity. This includes moving beyond cross-sectional score comparisons toward approaches that examine learning processes, exposure conditions, and developmental change over time. Without such designs, it is difficult to determine whether assessment outcomes reflect underlying reading difficulty or contextual constraints on language learning. Therefore, it is important to discuss the Limitations of Norm Referenced Assessment Research which often occupies a central place in SoR-aligned assessment frameworks. These tools are designed to compare individual learners' performance against population norms, often to identify those at risk of reading difficulty. While norm-referenced designs can be useful for large-scale screening, their validity in TESOL contexts is inherently limited (Alzghoul et.al, 2025).

Normative samples typically reflect monolingual language development patterns and assume relatively homogeneous exposure to English. In multilingual contexts, however, learners' performance distributions are shaped by heterogeneous exposure histories. As a result, deviations from monolingual norms may reflect opportunity gaps rather than cognitive or linguistic deficits (Messick, 1995).

From a methodological standpoint, this raises questions about the appropriateness of applying fixed cut scores across linguistically diverse populations. Studies that rely exclusively on norm-referenced outcomes risk conflating difference with disorder, particularly when they fail to control for exposure variables. For EFL learners, norm-referenced classification may therefore lack construct validity, even when instruments demonstrate strong psychometric properties in monolingual samples.

To address these limitations, researchers must incorporate analytic strategies that disaggregate exposure effects from skill-related variance. This may involve stratifying samples by instructional context, controlling for input frequency, or employing within-subject designs that track growth over time. Without such controls, assessment research risks reinforcing deficit-oriented interpretations.

Longitudinal research designs are particularly well suited to examining reading development in EFL contexts. By tracking learners across time, researchers can distinguish between temporary performance differences associated with limited exposure and persistent difficulties indicating reading-related disorders. Such designs align closely with the developmental sensitivity emphasized in linguistically responsive assessment frameworks.

Growth-oriented analyses also allow researchers to examine responsiveness to instruction, a critical indicator of learning potential. Learners who demonstrate accelerated growth following increased exposure or targeted instruction may not require intensive remediation, even if their initial performance falls below normative benchmarks. In contrast, learners whose growth remains limited despite adequate exposure may warrant further investigation (Lantolf & Poehner, 2014).

Despite their value, longitudinal studies remain underrepresented in research on SoR-aligned assessment in TESOL contexts. This gap reflects both logistical challenges and systemic preferences for short-term outcome measures. Addressing this imbalance is essential for developing assessment practices that are developmentally and ethically sound.

In addition to longitudinal designs, mixed methods approaches offer significant potential for advancing understanding of assessment use in multilingual contexts. Quantitative data can identify patterns of performance and growth, while qualitative data illuminate how assessment results are interpreted and enacted in classrooms.

Teacher interviews, classroom observations, and case studies provide insight into how assessment data are translated into instructional decisions. Such methods are particularly valuable for examining the role of professional judgment, an element that is central to linguistically responsive assessment but difficult to capture through standardized measures alone (Snow, 2002).

Mixed methods research also enables examination of unintended consequences of assessment practices, such as labeling effects, changes in instructional focus, or shifts in learner identity. These consequences are central to consequential validity and must be considered when evaluating the appropriateness of assessment systems in TESOL contexts (Messick, 1995).

Research on assessment in multilingual contexts carries ethical responsibilities that extend beyond methodological rigor. Researchers must remain attentive to how their findings are likely to be interpreted and used by policymakers and practitioners. Studies that emphasize deficits without contextual qualification risk reinforcing harmful narratives about multilingual learners.

Ethically responsible research should therefore foreground limitations, contextual variables, and alternative interpretations. In the context of SoR-aligned assessment, this means explicitly acknowledging the role of exposure, instructional opportunity, and

cross-linguistic influence. It also requires resisting simplistic conclusions that equate assessment performance with learner ability. By adopting ethically informed methodological practices, researchers can contribute to assessment frameworks that support equity rather than exclusion. One of the least examined dimensions of Science of Reading-aligned assessment reform is the role of accountability systems in shaping how assessment data are interpreted and used. While SoR frameworks are often presented as neutral applications of cognitive science, their enactment occurs within policy environments characterized by performance monitoring, benchmarking, and institutional evaluation. In such contexts, assessment data are rarely used solely for diagnostic purposes; they are also mobilized to demonstrate compliance, effectiveness, and reform success.

In many Middle Eastern education systems, accountability structures place considerable emphasis on measurable outcomes, particularly in English. School performance indicators, inspection reports, and accreditation reviews frequently rely on standardized assessment data as evidence of instructional quality. Under these conditions, SoR-aligned assessments may be used less as tools for understanding learner development and more as instruments for institutional comparison. This shift in purpose has significant implications for validity.

When assessment outcomes are tied to accountability consequences, interpretive flexibility diminishes. Teachers and school leaders may feel pressure to adhere strictly to benchmarks and cut scores, even when those benchmarks are misaligned with learners' linguistic realities. As a result, assessment interpretation becomes increasingly standardized, while contextual nuance is treated as a threat to comparability rather than a requirement for validity (Messick, 1995).

This narrowing of interpretation disproportionately affects EFL learners. Because their performance often reflects exposure variability, they are more likely to fall below monolingual norms. In accountability-driven systems, such outcomes may be interpreted as evidence of instructional failure or learner deficiency, prompting intensified intervention or curricular restriction. Over time, this dynamic can lead to systematic misrepresentation of multilingual learners' abilities.

Debates surrounding the Science of Reading are frequently framed in terms of evidence hierarchies. Randomized controlled trials, large-scale quantitative studies, and psychometric validation research are often positioned as the highest forms of evidence, while qualitative, interpretive, or context-sensitive research is treated as secondary. While rigorous quantitative research has made invaluable contributions to understanding reading development, its dominance has unintended consequences for assessment reform in multilingual contexts.

In TESOL settings, many of the most consequential variables shaping literacy developments such as language exposure, instructional mediation, and sociolinguistic environment—are difficult to capture through standardized measures alone. Teacher observations, classroom discourse analysis, and learner histories provide critical insights into how reading skills develop in context. When such forms of knowledge are marginalized, assessment systems risk privileging decontextualized scores over meaningful interpretation (Snow, 2002).

This hierarchy of evidence also shapes professional authority. Teachers' experiential and interpretive knowledge is often discounted in favor of externally produced metrics, reinforcing a technocratic model of assessment. In multilingual classrooms, where standardized tools are least informative in isolation, this dynamic undermines responsible decision-making. Linguistically responsive assessment requires not only better tools, but also a broader conception of what constitutes legitimate evidence.

9. The Risk of Reifying “Reading Difficulty” in Multilingual Learners

Another critical consequence of unadapted SoR-aligned assessment is the reification of reading difficulty as a stable learner trait. When assessment outcomes are interpreted without attention to language development, learners may be labeled as struggling readers based on performance that reflects temporary or contextual factors. These labels, once applied, can be difficult to reverse.

Research on labeling effects suggests that early classification as a low performer can shape teacher expectations, instructional opportunities, and learner self-concept. In multilingual contexts, where language proficiency is still developing, premature labeling risks pathologizing normal development (Cummins, 2000). SoR-aligned assessments, when misused, may therefore contribute to the construction of reading difficulty rather than its accurate identification.

This risk is amplified when assessment data are used to justify long-term placement decisions, such as streaming or withdrawal from mainstream instruction. Learners who are repeatedly assessed against monolingual benchmarks may internalize deficit narratives, even as their language proficiency improves. From an ethical perspective, such practices raise serious concerns about fairness and proportionality.

A linguistically responsive approach challenges the notion that reading difficulty can be determined solely through early performance on decontextualized tasks. Instead, it emphasizes developmental trajectories, responsiveness to instruction, and opportunity to learn as central considerations in identification. A persistent tension in discussions of SoR-aligned assessment lies between the pursuit of scientific rigor and the demand for educational justice. Proponents of standardized assessment often argue that uniform measurement ensures fairness and objectivity. However, in multilingual contexts, uniform measurement may produce systematically unequal outcomes.

Educational justice requires acknowledging that learners enter classrooms with unequal access to linguistic resources. Assessment systems that ignore this reality risk perpetuating inequality under the guise of neutrality. From this perspective, fairness is not

achieved by treating all learners identically, but by interpreting assessment outcomes in ways that account for differential opportunities. (Snow, 2002).

Reconciling rigor with justice does not require abandoning scientific principles. Rather, it involves expanding the evidentiary base upon which assessment decisions are made. Linguistically responsive assessment integrates quantitative data with qualitative insight, static measures with dynamic evidence, and standardized tools with professional judgment. This integration enhances, rather than diminishes, validity. Finally, it is important to emphasize that linguistically responsive assessment should not be construed as resistance to the Science of Reading. On the contrary, it represents a necessary evolution of SoR principles as they are applied in linguistically diverse contexts. By addressing limitations in assessment interpretation, linguistically responsive approaches strengthen the scientific and ethical foundations of literacy reform.

In the Middle East, where educational systems are navigating the intersection of global reform agendas and local linguistic realities, such evolution is essential. Assessment practices that align with both reading science and second language development are better positioned to support learners equitably and sustainably. By reframing assessment as an interpretive, context-sensitive practice rather than a purely technical procedure, this article advances a model of reform that honors both empirical evidence and educational responsibility.

10. Directions for Future Research

While this article has focused on conceptual and theoretical considerations, there is a clear need for further empirical research to support linguistically responsive SoR-aligned assessment in TESOL contexts. One priority area is the validation of assessment tools adapted specifically for Arabic-speaking EFL learners. Studies examining construct validity, reliability, and predictive power across diverse educational settings would provide valuable evidence to inform practice.

Longitudinal research tracking learners' reading development over time is also needed. Existing research suggests that reading comprehension development in EFL/ESL learners may be different from monolingual English speakers. This is especially true in decoding and language comprehension over time. However, research on Arabic-speaking EFL contexts is limited. Al-Seghayer (2014) identifies several systemic constraints affecting English language teaching in Saudi Arabia that may influence exposure to English and classroom learning conditions. Saiegh-Haddad (2018) and Vettori et al. (2023) highlight the influence of linguistic features like diglossia and orthographic complexity on reading development. Such studies could help distinguish between temporary language-related challenges and persistent reading difficulties, informing more accurate identification and intervention decisions. Comparative research across Middle Eastern contexts would further illuminate how policy environments, school sectors, and exposure patterns shape assessment outcomes.

Finally, future research should explore the integration of dynamic assessment within SoR-aligned frameworks. Investigating how responsiveness to instruction can be systematically incorporated into assessment systems has the potential to reduce misclassification and support more equitable literacy education for multilingual learners.

11. Conclusion

The expansion of the Science of Reading into TESOL contexts offers significant promise for improving literacy outcomes. However, the effectiveness and ethical integrity of these reforms depend not only on instructional alignment, but also on assessment practices that accurately reflect learners' abilities and developmental trajectories.

In the Middle East, English literacy development is shaped by Arabic diglossia, cross-linguistic influence, and unequal access to English exposure. When SoR-aligned assessments are applied without sufficient adaptation or contextual interpretation, they risk conflating linguistic difference with reading difficulty.

Such practices undermine both assessment validity and educational equity.

This article has argued for a linguistically responsive orientation to SoR-aligned assessment—one that preserves the diagnostic strengths of reading science while accounting for the realities of second-language development. By integrating insights from reading science, bilingual literacy research, and contemporary validity theory, educators and policymakers can move toward assessment practices that are both scientifically grounded and ethically responsible. As SoR-informed reforms continue to shape literacy education globally, careful attention to assessment interpretation and use will be essential for ensuring that evidence-based practice serves all learners equitably.

Acknowledgments

Not applicable.

Author contributions

Dr. Salam Sammour contributed to the conceptualization of the study, prepared the original draft of the manuscript. Dr. Mohammed Al-Badawi and Dr. Ala Al-Tarawneh provided subject-matter expertise, contributed to the refinement of the manuscript through critical review and editing, and offered intellectual feedback throughout the writing process. Ala Yaghi contributed to the critical review and editing of the manuscript. All authors read and approved the final version of the manuscript.

Funding

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Abu-Rabia, S. (2019). The role of short vowels in reading Arabic: A critical literature review. *Journal of Psycholinguistic Research*, 48(4), 785-795. <https://doi.org/10.1007/s10936-019-09631-4>
- Al-Qahtani, A. A. (2016). Why do Saudi EFL readers exhibit poor reading abilities? *English Language and Literature Studies*, 6(1), 1-15. <https://doi.org/10.5539/ells.v6n1p1>
- Al-Seghayer, K. (2014). The four most common constraints affecting English teaching in Saudi Arabia. *International Journal of English Linguistics*, 4(5), 17-26. <https://doi.org/10.5539/ijel.v4n5p17>
- Al-Seghayer, K. (2019). Unique challenges Saudi EFL learners face. *Studies in English Language Teaching*, 7(4), 490-499. <https://doi.org/10.22158/selt.v7n4p490>
- Alzghoul, M. A., Abualasal, J., Alazzam, T. S., Al-Ghammaz, S. A. D., & Alzghoul, R. M. (2025). The Status Quo of Translation Technology Tools in Translator Training Programs at Jordanian Universities. *World Journal of English Language*, 15(7), 398-398. <https://doi.org/10.5430/wjel.v15n7p398>
- Bar-On, A., Shalhoub-Awwad, Y., & Tuma-Basila, R. I. (2018). Contribution of phonological and morphological information in reading Arabic: A developmental perspective. *Applied Psycholinguistics*, 39(6), 1253-1277. <https://doi.org/10.1017/S0142716418000310>
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5-51. <https://doi.org/10.1177/1529100618772271>
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Multilingual Matters. <https://doi.org/10.21832/9781853596773>
- Ehri, L. C. (2020). The science of learning to read words: A case for systematic phonics instruction.. *Reading Research Quarterly*, 55(S1), S45-S60. <https://doi.org/10.1002/rrq.334>
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93-99. <https://doi.org/10.1598/RRQ.41.1.4>
- Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties*. Wiley.
- Koda, K. (2005). *Insights into second language reading*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524841>
- Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education : Vygotskian praxis and the research/practice divide*. Routledge. <https://doi.org/10.4324/9780203813850>
- Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to

- teach English language learners. *Journal of Teacher Education*, 59(4), 361-373. <https://doi.org/10.1177/0022487108322110>
- Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 140(2), 409-433. <https://doi.org/10.1037/a0033890>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Perfetti, C. A. (2017). Lexical quality revisited. In E. Segers & P. van den Broek (Eds.), *Developmental perspectives in written language and literacy*, 51-67. <https://doi.org/10.1075/z.206.04per>
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer.
- Ritchey, K. D., Palombo, K., Silverman, R. D., & Speece, D. L. (2017). Effects of an informational text reading comprehension intervention for fifth-grade students. *Learning Disability Quarterly*, 40(2), 68-80. <https://doi.org/10.1177/0731948716682689>
- Saiegh-Haddad, E. (2018). MAWRID: A model of Arabic word reading in development. *Journal of Learning Disabilities*, 51(5), 454-462. <https://doi.org/10.1177/0022219417720460>
- Saiegh-Haddad, E., & Joshi, R. M. (2014). *Arabic literacy: The role of orthography and diglossia*. Springer.
- Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, 134(4), 584-615. <https://doi.org/10.1037/0033-2909.134.4.584>
- Snow, C. E. (2002). *Reading for Understanding: Toward an R&D program in reading comprehension*. RAND Corporation. Retrieved from http://www.rand.org/pubs/monograph_reports/MR1465.html
- Tunmer, W. E., & Hoover, W. A. (2019). The cognitive foundations of learning to read. *Australian Journal of Learning Difficulties*, 24(2), 75-93. <https://doi.org/10.1080/19404158.2019.1614081>
- Vaughn, S., Roberts, G. J., Miciak, J., Taylor, P., & Fletcher, J. M. (2019). Efficacy of a word- and text-based intervention for students with significant reading difficulties. *Journal of Learning Disabilities*, 52(1), 31-44. <https://doi.org/10.1177/0022219418775113>
- Vettori, G., Incognito, O., Bigozzi, L., & Pinto, G. (2023). Relationship between lexical, reading and spelling skills in bilingual language minority children and their monolingual peers. *Frontiers in Psychology*, 14, Article 1121505. <https://doi.org/10.3389/fpsyg.2023.1121505>