SAFARA

Revue internationale de langues, littératures et cultures

N°24
2026

Sommaire

The Effect of Task Instruction Language on Performance in English Reading Comprehension Assessment: An Experimental Study in Kebemer, Senegal

Mamadou Moustapha Sanghare

University of Sussex, UK

Résumé

Cette étude expérimentale examine l'impact de la langue des consignes sur la performance en compréhension écrite en anglais. 50 lycéens ont été répartis aléatoirement en deux groupes égaux : l'un a reçu les consignes en français, l'autre en anglais. Les résultats montrent que les élèves ayant reçu les consignes en français ont obtenu des scores nettement plus élevés (*d* de Cohen = 1,94). L'analyse qualitative montre que ces élèves ont mobilisé des stratégies de compréhension orientées vers le sens, alors que les élèves ayant reçu les consignes en anglais ont plutôt recouru à des stratégies superficielles, en raison d'une mauvaise interprétation des consignes et des questions. Ces résultats indiquent que les évaluations monolingues en anglais introduisent une variance non pertinente au construit et sous-représentent les capacités de compréhension.

Mots-clés : compréhension écrite ; langue des consignes ; langue d'évaluation ; validité du construit ; anglais langue étrangère ; apprenants sénégalais

Abstract

This experimental study investigated the impact of task instruction language on the performance of Senegalese secondary school students in English reading comprehension. Fifty students were randomly assigned to two groups: one received task instructions in French and the other in English. The results indicate that students receiving French task instructions scored significantly higher (Cohen's d = 1.94). Qualitative analysis demonstrated that those students used comprehension strategies that focused on meaning. Students who received instructions and questions in English employed surface-level strategies because they misunderstood the task. These findings suggest that monolingual English

═══

assessment introduces substantial construct-irrelevant variance and underrepresents students' reading comprehension abilities.

1. Introduction

Senegal is a multilingual country where French is the primary language of education and administration. National languages such as Wolof, Pulaar, and Sereer are used for everyday communication (Sanghare, 2025). English remains a foreign language taught in secondary education from the age of 12. In 2024, it was introduced into primary education. Exposure to it is limited to schools, with little presence in the linguistic landscape (Gorter, 2006) of Senegal. However, in reading comprehension tests, questions and instructions are usually written entirely in English. This monolingual format may not align with students' English proficiency levels. The first obstacle they have to overcome to demonstrate their comprehension of a text may not be linked to the text itself, but to the understanding of instructions and questions.

Preliminary classroom observations indicate that this leads many students to approach the challenge by using decoding strategies (Kane & Martin, 2024) rather than attempting to understand instructions. In such cases, success on reading comprehension tasks may result from word recognition rather than actual cognitive understanding. Thus, a reading comprehension assessment may not be construct-valid. This raises the general question: Is the language used in task instructions and questions a factor in Senegalese students' EFL reading comprehension? Specific operational questions include: Do students who receive test instructions in English perform differently from those who receive test instructions in French when working on reading comprehension assessments? What are the implications of any test score differences for the construct validity of English reading comprehension assessments in Senegal?

This study investigates the impact of the language of task instructions and questions on students' English reading comprehension performance in a public secondary school in Kebemer, Senegal. It analyses the effect of the

===

language of task instructions on students' responses to the reading comprehension questions. Finally, this study seeks to provide practical implications to improve the accuracy and validity of English reading comprehension assessments in similar Senegalese school contexts.

2. Reading comprehension and testing in EFL

Cognitive and linguistic dimensions of reading comprehension

Reading comprehension is a complex mental process that requires constructing meaning from written texts through the interplay of linguistic and background knowledge, as well as strategic processes. Urquhart and Weir (2016) define it as an interactive process between the reader and linguistic input to create meaning based on the reader's prior knowledge. In EFL contexts, comprehension requires understanding the linguistic, semantic, and cultural dimensions of English (Grabe, 2008).

There is a clear distinction between decoding and comprehension in the literature about psycholinguistic processes and reading pedagogy (Brunfaut, 2022; Gough & Tunmer, 1986; Hoover & Gough, 1990; M. B. Kane & Martin, 2024; Urquhart & Weir, 2016). Decoding is identifying and connecting written words to either their aural or visual form without a mental representation of their meaning (Hoover & Gough, 1990; Kane & Martin, 2024). It involves perception and mechanical processes, such as lexical recognition or visual matching. However, comprehension is a higher-order reasoning process. It requires making sense of content through mobilizing prior knowledge, making inferences, paraphrasing, and developing an overall meaning composition (Grabe, 2008; Kane & Martin, 2024; Reed, 2020).

Kintsch and Rawson (2005) introduce the Construction-Integration Model. In this framework, comprehension takes place when a reader constructs a mental representation of the text by integrating new information with prior knowledge. Without prior knowledge, comprehension does not occur. This is because the bottom-up processes of word recognition and syntactic understanding do not connect with the top-down processes of using, for

example, context and aiming to accomplish a purpose in reading (Urquhart & Weir, 2016).

Foreign language learners and reading comprehension

As foreign-language readers, learners in this study seem to read texts in English as literal translations. This relies on their literal understanding rather than inferential or critical comprehension (Kane & Martin, 2024). This tendency mirrors traditional pedagogies, in which reading instruction is often treated as a test-based skill rather than a communicative process (Grabe, 2008). Kane and Martin (2024) note that learners who read in a second language tend to have difficulty comprehending texts due to lower linguistic and cognitive resources than first-language speakers.

Evidence suggests that foreign-language readers typically show lower levels of grammatical knowledge. This creates some difficulties in processing sentence structures and relationships within a text (Alderson et al., 2016; Brunfaut, 2022; Spencer & Wagner, 2017). They are often slower because of lower levels of automatism in word recognition. This results in greater time and mental effort spent on decoding individual words rather than on meaning. As Brunfaut (2022) observes, L2 readers work with limited linguistic resources compared to L1 readers. In Senegal, where English functions as an L3, this constraint is further intensified when assessment instructions require L3 processing.

Studies on cognitive load suggest that linguistic processing demands an overwhelming amount of working memory capacity. Sweller (1988) outlines three types of cognitive load: intrinsic, which is the effort associated with task difficulty; extraneous, which is effort associated with how tasks are presented; and germane, which refers to processing that builds learning. For learners with limited language proficiency, there may be an increase in extraneous cognitive load created by ambiguous task instructions. Furthermore, foreign-language readers may lack sufficient background and cultural knowledge of the subject area or the discourse conventions in texts written in the target language (Alptekin, 2006). These challenges force them to rely on bottom-up processing rather than top-down processing. This makes

it challenging for L2 readers to perform well in reading comprehension compared to L1 readers (Kane & Martin, 2024).

Historically, the Direct Method and the Audio-lingual Method discouraged the use of learners' first language, encouraging the target language only in class (Sundari & Febriyanti, 2021). The argument was that providing exposure to the target language is necessary to maximise exposure. Despite these usual arguments for using the target language, Gündüz and Kılıçkaya (2021) conducted a study that shows that learners who receive task instructions in their L1 may better understand the task concept, focusing on the task content rather than the process of understanding the task instructions. This leads to better student performance. Students in the experimental groups received L1 task instructions and achieved higher scores in listening activities than those in the control group, which received L2 task instructions.

Testing and construct validity

The notion of construct validity in language testing refers to the extent to which a test assesses the ability it claims to evaluate and the degree to which test scores represent the intended construct and not unrelated factors (McNamara, 2010; Messick, 1987, 1995; Chapelle & Lee, 2021; Yao et al., 2022). Recent theoretical developments in language testing suggest that construct validity relies on ensuring test scores reflect the intended measures and that external conditions have no impact on test results (Chapelle & Lee, 2022). Kane's framework for establishing valid interpretations of test scores relies on coherent chains of inference from the observed performances of students to the constructs being measured (Kane, 2022). When linguistic barriers in instructions impede learners' ability to demonstrate their actual competence, the inferential chains are broken.

In EFL assessment, learners' limited proficiency, unfamiliarity with task types, and misunderstanding of instructions may constitute a challenge to construct validity (Alderson, 2005). Bachman (2010) extended this idea to communicative language testing, claiming that language ability should be seen in terms of grammatical knowledge and pragmatic competence. A valid

EFL assessment must measure not merely knowledge of language, but the ability to use language meaningfully in communicative contexts (McNamara, 2010). If a test does not include both components, it may instead measure qualities peripheral to proficiency, such as memory, guessing skills, or test-taking strategies, not language proficiency itself. Learners perform better when instructions are in a language they already know, as interpretation is more straightforward.

Messick (1987, 1995) indicates that two factors can undermine a test's construct validity: construct underrepresentation and construct-irrelevant variance. Construct underrepresentation occurs when the assessment does not adequately test the whole domain of knowledge, skills, and abilities that comprise the intended construct. Instead, the evaluation represents only a narrow portion of what it intends to measure. Construct-irrelevant variance occurs when performance on a test is influenced by factors external to the intended construct, introducing extraneous variance that contaminates score interpretation (Messick, 1995). As Brunfaut (2022) observes, L2 reading comprehension construct validity may be influenced by factors related to the texts, the questions, the test takers, etc.

Alderson (2005) stresses that reading comprehension tests may reflect the specificity of learners' background knowledge or the degree of difficulty of the task, rather than the relevant textual knowledge. For many EFL learners, misunderstanding a task's instructions and questions, particularly when those instructions are in English, may interfere with their performance (Bachman & Palmer, 2010). Learners may struggle to show authentic comprehension due to a misinterpretation of the task instructions. Thus, construct validity in EFL testing may be problematic in Senegal for a further reason. Reading comprehension task instructions are written in English, and students are to answer in English. This means that it is not only a reading comprehension test but also a writing competence test.

Pillay (2025) investigated how South African township schools conduct their assessments and found that allowing learners to utilise their full repertoire during tests significantly improved comprehension and assessment accuracy. Teachers involved in this study indicated that multilingual practices enabled

deeper understanding than when assessments were conducted only in English. This suggests that monolingual tests systematically under-assess learners' conceptual knowledge because they limit their ability to utilise the full range of language resources.

This challenges monolingual assessments' assumptions. When a learner's ability to create meaning exceeds their target language proficiency, a monolingual assessment may be testing the following constructs: comprehension, productive vocabulary, and strategic competence under linguistic constraints, rather than measuring comprehension alone. As Pillay (2025) and Gündüz & Kılıçkaya (2021) demonstrate, removing language barriers during assessments may provide a more accurate measure of comprehension. Therefore, the language in which instruction and questions are delivered is an essential component affecting construct validity in multilingual contexts.

3. Methods

Research design

The study used a mixed-methods experimental design. It combined quantitative analysis of test scores with qualitative examination of response patterns to understand both performance outcomes and cognitive strategies. The study involved an experimental group and a control group, which were compared to assess the effect of the language used in task instructions on English reading comprehension. In this research, the experimental group received task instructions in French and had to answer in English. In contrast, the control group received instructions and activities in English and had to answer in English.

In line with the objectives of the study and the experimental design described above, the hypotheses were formulated. H_0: There will be no statistically significant difference between students who receive task instructions and questions in French and those who receive them in English, concerning their test scores in reading comprehension. H_1: There will be a statistically

significant difference between the two groups, with higher scores for the first group.

Participants and sampling

This research was undertaken in one school in Kebemer in the Louga region. The single-site design of this study creates an experimental environment that enables a rigorous examination of the causal relationship between the language used in task instructions and reading comprehension test scores. Thus, findings from this study were limited to this school. Tests were administered simultaneously to ensure identical conditions. Fifty secondary school students with similar levels were randomly divided into two groups of 25 each. Both groups received the same English language reading text on education.

The reading passage was taken from a standardised school examination. To eliminate any prior knowledge of the text as a bias, both groups were surveyed to determine whether they had been exposed to it. The text was approximately 250 words long. Each group was tested on 10 comprehension questions in three styles: 4 multiple-choice questions, three true-or-false questions with justification, and three open-ended questions. An example of an open-ended question was: "Why is improving access to secondary education important for girls?"

To ensure comparability between the groups, baseline equivalence was assessed prior to the intervention using available data: age, years of English language learning, and gender. Additionally, a reading comprehension test on international migration in Senegal was administered to confirm the homogeneity of their levels. One male student was excluded due to his very high grade (19/20), which contrasted with the overall grades. The pre-test's results (t=0.73, p=0.47, and d=0.21) revealed no significant differences in test scores between the experimental and control groups, indicating comparable starting points.

==

Variable	Experimental group	Control group	Statistic test	*p*-value	d
Age	M=18.7 (SD: 0.57)	M=18.58 (SD: 0.53)	t=0.774	0.44	0.22
Pre-test score	M=10.48 (SD: 1.42)	M=10.20 (SD: 1.21)	t=0.73	0.47	0.21
Gender	11F - 14M	12F- 13M	X^2=0.0804	0.77	--
Sample size	N=25	N=25	--	--	--

Table 1: Baseline characteristics by groups

Data analysis

An independent-samples t-test was used to compare learners' mean scores to determine whether there was a significant difference between the two groups. It examined whether the language of instruction had a significant effect on students' reading comprehension. The t-test produced a *t*-value and a *p*-value. The effect size (*d*) was calculated to determine the practical significance of the difference. It complements the *p*-value by quantifying the magnitude of difference beyond statistical significance. It was interpreted according to Cohen's (2009) conventions: an effect size of 0.2 is a small effect, 0.5 is a medium effect, and 0.8 is a significant effect.

To complement the quantitative analysis and explore how the language of instructions and questions influenced students' cognitive strategies, a qualitative study of student responses was conducted. The open-ended responses and justifications to the true-or-false questions were all read systematically to identify patterns in how participants approached the task. Answers were classified into the following categories: correct answer with comprehension, lexical matching, copying text, and incomplete or unrelated response. Patterns were compared to establish systematic differences in mental engagement.

4. Results and analysis

Statistical description and t-test results

Scores across both groups ranged from 06/20 to 16/20, but with a different distribution. In the experimental group, most students were in the mid- to high-score range. Most scores fell between 12 and 16; six students scored beyond 15. However, seven students scored below 12, yielding a mean score of 12.9, a median of 12, and a standard deviation of 2.17 for the experimental group. In contrast, the control group had lower scores. Many students scored between 05 and 09, lowering the overall mean. Most students scored between 06 and 10; five scored 12. Generally, the group has a mean score of 8.7, a median score of 08, and a standard deviation of 2.15. These results indicate that the experimental group had higher mean and median scores and outperformed the control group.

Both groups have nearly identical standard deviations, indicating that their score distributions are very similar around their means. This reveals that variability within each group is consistent, and neither group shows greater heterogeneity. Consequently, any observed differences between these group means are more likely to reflect actual effects rather than differences in score dispersion.
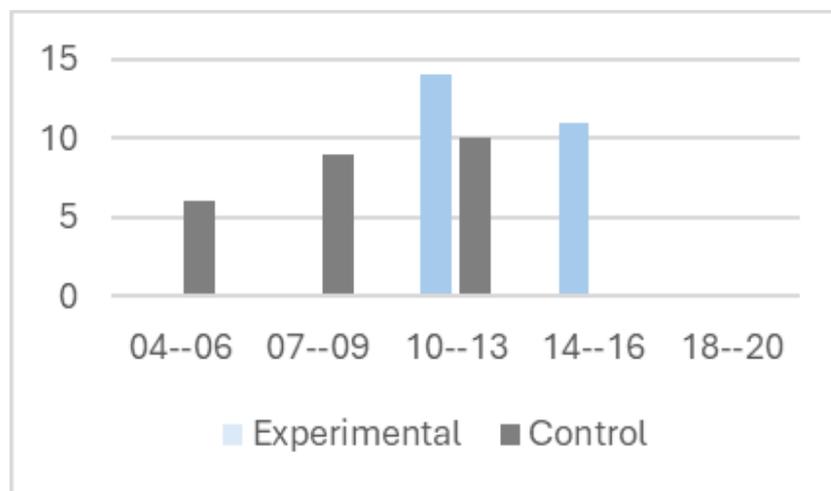
Figure 1. Distribution of scores by group

==
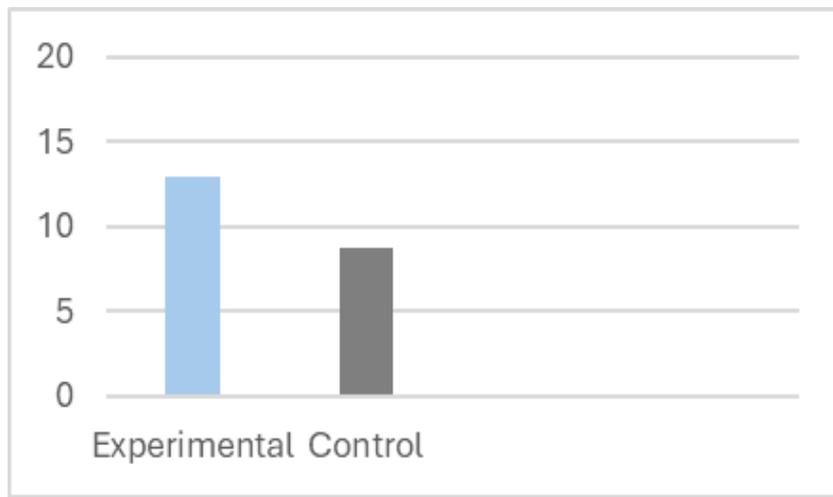
Figure 2. Mean scores by group

Group	n	mean	median	variance	std dev	min	max
Experimental	25	12.9	12	4.69	2.17	10	16
Control	25	8.7	08	4.62	2.15	06	12

Table 2. Description of statistics by group

The t-test results also confirmed the trend. The *t*-value obtained was 6.88. Since it is far greater than the critical value H$_0$ for α = 0.05, the difference between the two groups is very significant. The *p*-value was 0.0001, well below the 0.05 threshold. This demonstrates that the probability of observing the difference by chance alone is low. Thus, the null hypothesis H$_0$ is rejected. The value of Cohen's *d* was 1.94, which suggested a considerable effect size. Therefore, not only was there a statistically significant difference, but the size of the difference was large enough to be meaningful in practical and pedagogical terms. These findings support the main hypothesis H$_1$. Students

who received task instructions and questions in French performed significantly better than those who received them in English.

Analysis of response patterns

The qualitative analysis of students' responses revealed clear and systematic differences in the strategies used by the two groups. These patterns demonstrate how the language of task instructions shaped learners' cognitive engagement during the reading assessment. The students in the control group showed response patterns typical of readers under high linguistic and cognitive load.

In the control group, many students frequently used low-level strategies rather than engaging meaningfully with the intended meaning of the questions or the passage. A standard behaviour was lexical matching. Students identified sentences or sentence parts of the passage that contained the exact key words as the question. They selected these sentences as support, mistakenly believing they equaled a correct answer. For example, participants were asked whether the sentence "some children leave school without basic literacy or numeracy" is true or false, and to justify their answer by quoting the text. However, 21 students copied "education is not just about literacy and numeracy" from the text, showing they relied on finding similar words rather than engaging with the meaning.

Another observed behaviour was sentence copying, in which students reproduced entire lines from the text even when they did not respond to the question. For instance, when asked "How does girls' education contribute to a nation's development?", many students copied the sentence "It is important for a nation's development". However, the question required them to explain how girls' education contributes to society by describing benefits such as increased productivity and improved health, all of which are mentioned in the text.

Such approaches appear typical of foreign language readers who rely on bottom-up processing strategies rather than on top-down comprehension

==

(Alderson et al., 2016; Spencer and Wagner, 2017). In addition, the use of mechanical strategies also demonstrates that students did not build a mental representation of the text by connecting new information to prior knowledge, as outlined in Kintsch and Rawson's (2005) Construction-Integration Model. When students do not understand instructions, they cannot use the integrative processes required for truly understanding (Urquhart and Weir, 2016).

Students in the experimental group used instructions to navigate to specific information needed for their response, rather than surface-level techniques. Generally, students' responses were more complete, coherent, and concise, providing explicit evidence that they understood the question's underlying meaning. For example, when asked why improving girls' access to secondary education is essential, most students answered that educated mothers are more likely to ensure their children are vaccinated, receive proper nutrition and attend school.

The experimental group also showed greater ability to justify their true or false answers with objective evidence from the text. For example, when dealing with a statement that was initially written in French about gender parity in primary education having been fully achieved, most students who marked the statement as false justify their point by quoting the text "72 million children worldwide do not attend school, and 54% of the unschooled are girls". Others wrote "considerable gaps remain, particularly for girls". Most students in the control group could not provide a valid justification. Thus, the justifications used by students in the experimental group directly referenced the ideas presented in the passage, indicating that they were engaging with the meaning.

Removing the language barrier allowed students to allocate more cognitive resources to the passage. Rather than devoting time to decode the meaning of the question, students could focus on constructing meaning, selecting relevant information and using inference skills. The higher-quality responses indicate greater cognitive processing. These results provide evidence in support of Kintsch and Rawson's (2005) Construction-Integration Model, which states that comprehension occurs when new information from the text is integrated with prior knowledge. The use of French allowed students to use intellectual

resources to consolidate and process the new integration rather than to decode questions and instructions.

5. Discussion

Mechanisms and cognitive load

The effect size of 1.94 is considerable. Developing academic literacy in French reduces unnecessary cognitive load compared to English instruction conducted in the students' less familiar L3. Sweller's (1988) tripartite cognitive load model clearly explains how the control group had to process three types of information simultaneously: decoding vocabulary from the instructions, reading the syntax of the instructions, and processing the data to form a mental representation of what to do. These tasks consumed working memory before the reading even began. As a result, there was an overload of extraneous cognitive load while providing less than half the necessary capacity to process germane cognitive load.

In contrast, the experimental group only needed to process the information necessary to complete the task. Therefore, participants in this group could use all their working memory to create the required task representation. This may suggest that students processing instructions in English consume much working memory. This creates a significant cognitive load on a task designed to assess reading comprehension rather than instruction comprehension.

When task instructions and questions are delivered in a language learners are academically proficient in, they will have a better understanding of the task expectations. They will also display an ability to focus on the content of the passage and to demonstrate accurate understanding. As a result, this reduces threats to validity and the possibility that tests outcome reflects peripheral skills, such as word recognition or literal copying, rather than comprehension. These cognitive mechanisms manifest in observable differences in student task completion strategies.

==

Cognitive strategies and construct invalidity

Results are consistent with the Construction-Integration Model (Kintsch & Rawson 2005). When students cannot construct appropriate situation models due to language-related obstacles, they will be unable to integrate information from the passage to meet the task's demands. Qualitative evidence indicates that students in the control group produced text-based representations of the passage based solely on surface-level features. Therefore, they did not produce situation models. Participants in the experimental group were able to produce situation models and thus achieve a proper integration between the task requirements and the passage's meaning.

The strategic pattern shown represents evidence of the construct underrepresentation, as defined by Messick (1987, 1995). They do not provide adequate proof of how well students understand what they have read. When students struggle to interpret a question, they cannot demonstrate their full range of reading abilities. As the results indicate only surface-level knowledge of the text and do not reflect their full ability to read and understand it, the assessment covers only a portion of the reading comprehension process. Therefore, the evaluation does not provide a thorough representation of critical reading comprehension dimensions such as synthesising, analysing and making inferences from the text. As a result, the assessment does not accurately measure students' reading comprehension competence, leading to construct underrepresentation and invalidity.

This research also provides evidence of Messick's (1995) construct-irrelevant variance. Indeed, the use of English instructions in the reading comprehension assessment creates construct-irrelevant variance because they require greater linguistic processing to understand. Test scores reflect how comprehension processes are overwhelmed with construct-irrelevant demands. The strategic differences identified raise serious questions about the appropriateness of current testing methods in Senegal and other multilingual contexts.

The surface-level strategies shown by students in the control group demonstrate that the linguistic barrier imposed a systematic constraint on

cognitive processing. They also indicated that participants in the control group exhibited what Urquhart and Weir (2016) call expeditious reading, which is about choosing strategies based on task constraints rather than comprehension goals.

Implications for assessment policy and pedagogical practice

This research has broader implications for the validity of assessments in multilingual contexts. Although conducted in Senegal, the linguistic situation analysed in this research is representative of educational contexts across postcolonial countries. The large effect size suggests that when students develop academic literacy skills in their L2, delivering instructions for assessments in that L2 rather than the L3 significantly reduces construct-irrelevant variance, thereby improving test validity.

This aligns with Pillay (2025), who demonstrated that removing linguistic barriers improves assessment accuracy. While Pillay showed this using learners' linguistic repertoire at the time of assessment, this current study provides evidence that this principle also applies when using an academic second language. The benefit of both methods is that they eliminate the conflation of language proficiency with the target construct. The result also extends Gündüz and Kılıçkaya's (2021) findings by demonstrating that not only L1, but also an academic L2, can reduce linguistic barriers in assessment when the L1 is not available as an educational language. Together, the results of these studies show that it is necessary to provide instructions in a language that a student has sufficient proficiency to understand the task requirements, thereby reducing cognitive load.

These results demonstrate that construct validity is a significant concern in the design of EFL reading tests. Systematic variation in cognitive strategy use across groups provides substantial evidence of construct under-representation (Bachman & Palmer, 2010). When students use lexical matching rather than comprehending the text, the assessment of reading comprehension fails, as it becomes a measure of a different construct altogether. When test scores reflect attributes that are not representative of the intended construct, as noted

p. 204

by McNamara (2010) and Yao et al. (2022), this consequence undermines test validity.

In Senegal, where national English exams are conducted entirely in English, there are inherent biases in the tests. Language assessment studies recognise how much the social and political context impacts the design of tests (Young, 2022). In postcolonial contexts, monolingual English assessments reflect inherited colonial language hierarchies rather than actual educational needs. The principles of Universal Design in test design suggest that providing instructions in learners' stronger academic language is not an unfair advantage but rather the removal of construct-irrelevant barriers (Abedi, 2022). Thus, the design of national tests in Senegal appears to introduce construct-irrelevant variance in reading comprehension across two components: comprehension of written English texts and understanding of English-language instructions. When responses must be in English, a third consequential component of written English production is required, leading to a triple conflation of three different constructs: reading comprehension, instruction processing, and writing.

Thus, assessment reforms are needed. In the short term, national exams should include French translations of task instructions while preserving all English content in the reading passages. In the mid-term, test designers should create assessment items that assess reading comprehension independently from writing. In the long term, research should determine whether oral response formats yield more valid measures of comprehension in L3 contexts than written assessment formats.

Limitations and future directions

Several limitations must be considered. First, our design was limited to a single research site and therefore cannot be generalised statistically. However, the large effect size (d=1.94) implies that the finding is also robust and may replicate in similar environments. Secondly, our assessment method required participants to respond in written English, which may have confounded the constructs of reading and writing.

Although this is indicative of authentic assessment practices in Senegal and thus contributes to ecological validity, future research should investigate whether similar effects can be obtained in different proficiency levels. Lastly, a single reading passage was employed that focused exclusively on education, raising concerns about generalizability across various types of texts and fields of content. Nonetheless, the identified cognitive mechanisms, higher extraneous cognitive load and differences in strategic processing should hold, irrespective of the content.

Future research should replicate this study by examining more than one Francophone country in Africa, such as Mali, Burkina Faso, Côte d'Ivoire, and Niger, as they share similar linguistic contexts. By comparing L1, L2, and L3 within the same country, it will be possible to determine whether students assessed at the L2 level have the same academic progress as those assessed at the L1 level, or whether proficiency levels limit their progress. Examining these issues longitudinally to determine how language may affect student test scores across different proficiency levels will enable us to develop appropriate assessment policies.

Conclusion

This experimental study examined the impact of the language of task instruction and questions on learners' performance in English reading comprehension in a Senegalese secondary school. The results provide strong evidence that the language used in instructions and questions significantly impacts students' test scores. Learners who received task instructions and questions in French outperformed learners who received them in English. They had higher mean scores, more complete and relevant responses, and greater success with literal and inferential questions. Qualitative analysis indicated that French instructions and questions allowed students to allocate cognitive processing resources to the text rather than to understanding the task requirements. Students receiving English instruction often used surface strategies such as lexical matching and copying sentences.

These illustrate the critical role that task instruction language plays in establishing construct validity for reading comprehension assessments in EFL contexts. Current testing practices in Senegal, which require English-only instruction, result in the underrepresentation of the students' reading comprehension abilities. This study reaffirms the importance of supporting learners through bilingual or scaffolded initial instruction when cognitive overload is an issue. This makes tests more accessible and engaging.

References

- Abedi, Jamal. "Accommodations and Universal Design." *The Routledge Handbook of Language Testing*, by Glenn Fulcher and Luke Harding, 2nd edn, Routledge, 2022, pp. 306–21. *DOI.org (Crossref)*, https://doi.org/10.4324/9781003220756-24.

- Alderson, J. Charles, et al. "Characteristics of Weak and Strong Readers in a Foreign Language." *The Modern Language Journal*, vol. 100, no. 4, Dec. 2016, pp. 853–79. *DOI.org (Crossref)*, https://doi.org/10.1111/modl.12367.

- Alderson, John Charles. *Assessing Reading*. 1. publ., 5. pr, Cambridge University Press, 2005. Cambridge Language Assessment Series. *K10plus ISBN*.

- Alptekin, Cem. "Cultural Familiarity in Inferential and Literal Comprehension in L2 Reading." *System*, vol. 34, no. 4, Dec. 2006, pp. 494–508. *DOI.org (Crossref)*, https://doi.org/10.1016/j.system.2006.05.003.

- Bachman, Lyle F. *Fundamental Considerations in Language Testing*. 1. publ., 11. [print.], Oxford University Press, 2010. Oxford Applied Linguistics. *K10plus ISBN*.

- Bachman, Lyle F., and Adrian S. Palmer. *Language Testing in Practice: Designing and Developing Useful Languages Tests*. 1. publ., 13. [impr.], Oxford Univ. Press, 2010. Oxford Applied Linguistics. *K10plus ISBN*.

==

- Brunfaut, Tineke. "Assessing Reading." *The Routledge Handbook of Language Testing*, by Glenn Fulcher and Luke Harding, 2nd edn, Routledge, 2022, pp. 254–67. *DOI.org (Crossref)*, https://doi.org/10.4324/9781003220756-20.

- Chapelle, Carol A., and Hye-won Lee. "Conceptions of Validity." *The Routledge Handbook of Language Testing*, by Glenn Fulcher and Luke Harding, 2nd edn, Routledge, 2022, pp. 17–31. *DOI.org (Crossref)*, https://doi.org/10.4324/9781003220756-3.

- Gorter, Durk. "Introduction: The Study of the Linguistic Landscape as a New Approach to Multilingualism." *International Journal of Multilingualism*, vol. 3, no. 1, Apr. 2006, pp. 1–6. *DOI.org (Crossref)*, https://doi.org/10.1080/14790710608668382.

- Gough, Philip B., and William E. Tunmer. "Decoding, Reading, and Reading Disability." *Remedial and Special Education*, vol. 7, no. 1, Jan. 1986, pp. 6–10. *DOI.org (Crossref)*, https://doi.org/10.1177/074193258600700104.

- Grabe, William. *Reading in a Second Language: Moving from Theory to Practice*. 1st edn, Cambridge University Press, 2008. *DOI.org (Crossref)*, https://doi.org/10.1017/CBO9781139150484.

- Gündüz, Tarkan, and Ferit Kılıçkaya. "The Effects of Instructions in L1 and L2 in EFL Listening Classes on A2 Level Learners." *Lublin Studies in Modern Languages and Literature*, vol. 45, no. 3, Oct. 2021, pp. 97–107. *DOI.org (Crossref)*, https://doi.org/10.17951/lsmll.2021.45.3.97-107.

- Hoover, Wesley A., and Philip B. Gough. "The Simple View of Reading." *Reading and Writing*, vol. 2, no. 2, June 1990, pp. 127–60. *DOI.org (Crossref)*, https://doi.org/10.1007/BF00401799.

- Kane, Michael T. "Articulating a Validity Argument." *The Routledge Handbook of Language Testing*, by Glenn Fulcher and Luke Harding,

2nd edn, Routledge, 2022, pp. 32–47. *DOI.org (Crossref)*, https://doi.org/10.4324/9781003220756-4.

- Kane, Mohamadou Bachir, and Katherine I. Martin. "Teaching EFL Reading in Senegal: Current Practices and Recommendations." *Southern African Linguistics and Applied Language Studies*, vol. 42, no. 2, Apr. 2024, pp. 164–85. *DOI.org (Crossref)*, https://doi.org/10.2989/16073614.2023.2226705.

- Kintsch, Walter, and Katherine A. Rawson. "Comprehension." *The Science of Reading: A Handbook*, edited by Margaret J. Snowling and Charles Hulme, Blackwell Publishing Ltd, 2005, pp. 211–26. *DOI.org (Crossref)*, https://doi.org/10.1002/9780470757642.ch12.

- McNamara, T. F. *Language Testing*. 8. [print.], Oxford Univ. Press, 2010. Oxford Introductions to Language Study. *K10plus ISBN*.

- Messick, Samuel. "Validity." *ETS Research Report Series*, vol. 1987, no. 2, Dec. 1987. *DOI.org (Crossref)*, https://doi.org/10.1002/j.2330-8516.1987.tb00244.x.

- ———. "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning." *American Psychologist*, vol. 50, no. 9, Sept. 1995, pp. 741–49. *DOI.org (Crossref)*, https://doi.org/10.1037/0003-066X.50.9.741.

- Pillay, Preya. "Code-Switching and Translanguaging in Formative Assessment: Supporting ELT Development." *Journal of Languages and Language Teaching*, vol. 13, no. 3, July 2025, p. 1256. *DOI.org (Crossref)*, https://doi.org/10.33394/jollt.v13i3.13080.

- Reed, Stephen K. "Comprehension." *Cognitive Skills You Need for the 21st Century*, by Stephen K. Reed, 1st edn, Oxford University Press, 2020, pp. 3–14. *DOI.org (Crossref)*, https://doi.org/10.1093/oso/9780197529003.003.0001.

- Sanghare, Mamadou Moustapha. "Sustaining the Empire's Legacies: Schooling and The Perpetuation of Colonial Subjectivity in Post-Colonial Senegal." *Safara*, no. 23, Jan. 2025, pp. 107–29. *DOI.org (Datacite)*, https://doi.org/10.5281/ZENODO.18059929.

- Spencer, Mercedes, and Richard K. Wagner. "The Comprehension Problems for Second-language Learners with Poor Reading Comprehension despite Adequate Decoding: A Meta-analysis." *Journal of Research in Reading*, vol. 40, no. 2, May 2017, pp. 199–217. *DOI.org (Crossref)*, https://doi.org/10.1111/1467-9817.12080.

- Sundari, Hanna, and Rina Husnaini Febriyanti. "The Use of First Language (L1) in EFL Classrooms: Teachers' Practices and Perspectives." *International Journal of Education*, vol. 14, no. 1, Apr. 2021, pp. 70–77. *DOI.org (Crossref)*, https://doi.org/10.17509/ije.v14i1.26375.

- Sweller, John. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science*, vol. 12, no. 2, Apr. 1988, pp. 257–85. *DOI.org (Crossref)*, https://doi.org/10.1207/s15516709cog1202_4.

- Urquhart, A. H., and Cyril J. Weir. *Reading in a Second Language: Process, Product and Practice*. First issued in hardback, Routledge, 2016. Applied Linguistics and Language Study. *K10plus ISBN*.

- Yao, Don, et al. "Investigating the Construct Validity of an ESL Test for Young Learners." *World Journal of English Language*, vol. 12, no. 5, May 2022, p. 59. *DOI.org (Crossref)*, https://doi.org/10.5430/wjel.v12n5p59.

- Young, Richard F. "Social Dimensions of Language Testing." *The Routledge Handbook of Language Testing*, by Glenn Fulcher and Luke Harding, 2nd edn, Routledge, 2022, pp. 63–80. *DOI.org (Crossref)*, https://doi.org/10.4324/9781003220756-7.