LANGUAGE BARRIERS AND ASSESSMENT VALIDITY IN MOROCCAN SCIENCE CLASSROOMS: EVIDENCE FROM DIAGNOSTIC TESTING

Zineb Labbar

Faculty of Science Education, University of Mohamed V, Rabat, Morocco
zineblabbar710@gmail.com

Abdellatif El Moudni

Faculty of Science Education, University of Mohamed V, Rabat, Morocco
zineblabbar710@gmail.com

Abstract

While Arabic is still the primary language of instruction and comprehension in Moroccan public schools, teachers frequently use French to evaluate their students. This study investigates the effects of language barriers on students' performance and the reliability of science tests. A sample of secondary students was given a diagnostic test in two different formats: (1) a bilingual version with Arabic translations of important terms and instructions, and (2) a French-only version. Examining the relationship between linguistic accessibility and test comprehension, response accuracy, and overall achievement scores was the goal. According to preliminary findings, students perform better when given Arabic translations, which raises the possibility that linguistic considerations rather than subject-matter expertise could account for some of the low achievement in science courses. Correlations between test versions, item difficulty, and student language

background will be further investigated through data analysis. By emphasizing the significance of language-inclusive testing policies in Morocco and comparable educational systems, this study seeks to inform fair assessment practices in multilingual contexts. By emphasizing the significance of language-inclusive testing policies in Morocco and comparable educational systems, this study seeks to inform fair assessment practices in multilingual contexts.

Keywords:

Assessment Validity, Language Barriers, Bilingual Testing, Moroccan Education, Science Achievement

Introduction

Language is critical in education and evaluation, particularly in multilingual educational settings where the instructional language varies from the students' native language. In Morocco, French serves as the language of instruction for many scientific subjects, despite most students being educated mainly in Arabic. This language environment poses specific difficulties for teaching and evaluation, as learners need to understand intricate scientific ideas in a language they have not completely mastered. Diagnostic evaluation, aimed at recognizing students' abilities and educational requirements, is especially attentive to these language obstacles. When students struggle to understand the meaning of test questions, their results may indicate language challenges instead of a deficiency in scientific knowledge. Despite continued reforms such as the 2019 Loi-Cadre 51-17 encouraging plurilingualism numerous Moroccan students still struggle in science, prompting concerns regarding the linguistic fairness of existing evaluation methods.

1. Literature Review

The notion of diagnostic assessment originates in the early 1970s within the formative and summative evaluation framework proposed by Bloom, Hastings, and Madaus (1971). Their *Handbook on Formative and Summative Evaluation of Student Learning* reconceptualized evaluation as a means to enhance instruction rather than merely classify students. The authors argued that evaluation should "enable teachers to make marked improvements in their students' learning" (p. v), emphasizing its role as an evidence-based tool for adjusting teaching and identifying learner needs what would later be termed *diagnostic assessment*. Following its conceptual emergence within Bloom's formative-summative framework in the early 1970s, diagnostic assessment evolved from a theoretical idea into a structured practice adopted in educational reforms worldwide. During the 1980s, evaluation scholars such as Scriven (1983) and Black and Wiliam (1986) expanded Bloom's view, identifying assessment as a continuous process for teachers' improvement and student learning rather than a snapshot fixed in time. Naturally, in the 1990s, diagnostic assessment began to feature explicitly in national curriculum reforms particularly within OECD and UNESCO member states highlighting learner-centered education and the identification of prior knowledge as prerequisites for equity and inclusion within "education for all." By the early 2000s, the notion involved even more and had been institutionalized within global frameworks: the OECD's *Formative Assessment* report (2005) and

UNESCO's Education for All (2000-2015) agenda both framed diagnostic assessment as a tool for addressing learning disparities and supporting personalized instruction. The focus gradually expanded from identifying cognitive gaps to encompassing socio linguistic and contextual variables influencing performance aiming its improvement. This evolution reflects a broader paradigm shift from standardized, summative testing toward diagnostic-formative systems designed to improve student outcomes and learning.

However, this process often faces multiple challenges. Fadlillah and Kusaeri (2024) highlight that several factors inhibit the effective implementation of diagnostic assessment, such as teachers' resistance to new methods, their limited understanding and lack of professional training, and insufficient resources and infrastructure. This finding aligns with broader empirical research, where similar constraints have been documented by Tang and Zhan (2021), confirming that contextual and systemic factors continue to hinder the practical realization of diagnostic assessment in classrooms. This global evolution also resonates in Morocco. In a diverse and delicate Moroccan educational context, *assessment* and *education quality have* been among the central anchors of every major educational reform adopted since the 1999 Charte Nationale d'Éducation et de Formation (CNEF). Each successive reform from the Programme d'Urgence (2009–2012) to the Vision Stratégique 2015–2030 and the Loi-Cadre 51-17 (2019) has reiterated the need to strengthen both assessment quality and linguistic equity as mechanisms for improving learning outcomes and ensuring social inclusion (CSEFRS, 2019; OECD/UNESCO,2021).

After independence (1956), the reinstatement of Arabic as a language of instruction and assessment was inevitable; however, since the 1970s, a persistent linguistic divide has remained between higher education taught in French and Arabized primary and secondary schooling. This "fracture linguistique," as described by El Amrani (2013) and later echoed by Elmadhi and Boukhch (2025), is claimed to have contributed to college withdrawal a finding also reported by the Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique (Higher Council for Education, Training and Scientific Research, 2018) report. These students enter college with a noticeably low level of French proficiency, as argued by El Amrani (2013). This may, in fact, be one of the major factors behind student dropout, as many struggle to adapt to a new academic environment filled with unfamiliar concepts and yet another language they must learn. This issue continues to shape debates around equity and language policy in Moroccan education (Ismaili et al., 2024).

For decades, the Arabization of primary and secondary education remained largely intact until recent years, when Law 51-17 (Loi-Cadre) came to light, emphasizing the urgency of adopting a plurilingual and alternating-language approach in education, particularly for scientific subjects.

This shift, while intended to ensure better alignment between secondary and higher education both now taught in French has sparked a national debate regarding the efficiency and fairness of the reform Elbazini & Boumazzou (2024). Critics argue that despite the language change, Moroccan students' performance in scientific disciplines remains low in international assessments (PISA 2018, PISA 2022) raising questions about whether linguistic reform alone can address deeper pedagogical and systemic issues.

Such linguistic disparities also raise questions about the validity and fairness of classroom and diagnostic assessments administered in a language that many students are still acquiring. Therefore, this study sought to explore, to what extent do language barriers influence students' performance in diagnostic assessments, particularly in scientific subjects taught in French?

This study aimed to examine how the language of assessment influences the accuracy of diagnostic results and the interpretation of students' learning difficulties.

2. Methodology
2.1 Research Design
This study employed a quasi-experimental one group pretest–posttest design using a mixed-methods approach. The quantitative component involved two diagnostic assessments of the same scientific topic administered to the same group of middle school students: one bilingual (French including Arabic translation) and one monolingual (French only).

The qualitative component consisted of a focus group discussion conducted after the second test to explore students' perceptions of language-related challenges. This design allowed us to compare student performance across two language conditions and for the interpretation of results through students' self-reported experiences.

2.2 Participants

The participants were primary-level science students (n = 27) enrolled in a Moroccan public middle school. All students had received their primary instruction in Arabic and were currently learning scientific subjects taught in French, following the implementation of Loi-Cadre 51-17 (2019), which promotes bilingual and alternating-language instruction. The group was purposefully selected to represent learners transitioning between teaching in Arabic and in French education.

2.3 Instruments

Two diagnostic tests were used on the same scientific topic to ensure content comparability:

- Test 1 (Bilingual version): The test was written in French and accompanied by Arabic translations of key terms and instructions to support comprehension.
- Test 2 (Monolingual version): The same test items were presented entirely in French to replicate authentic assessment conditions under current science instruction.

Both instruments included multiple-choice and short-answer questions measuring conceptual understanding, reasoning, and use of science vocabulary. The tests were reviewed by two science teachers to ensure content equivalence and linguistic clarity.

2.4 Data Collection and Procedure

The two diagnostic tests were administered two weeks apart under identical classroom conditions. Students were not informed that both versions contained equivalent content to minimize expectancy effects. Responses were collected and scored using the same rubric to ensure consistency and comparability between the two administrations.

Following the second test, a formal focus group discussion was conducted with all participating students (n =6) to explore their perceptions of language-related difficulties encountered during the assessments. The session lasted approximately 20 minutes and followed a semi-structured format with open-ended prompts focusing on:

1. Difficulties in interpreting test questions written in French;
2. Whether language interfered with their ability to demonstrate scientific understanding;
3. Their comparative perceptions of the bilingual and monolingual versions.

The discussion was conducted in a relaxed classroom setting and facilitated by the researcher. Notes were taken during the session, and key statements were transcribed immediately afterward**.**

2.5 Data Analysis

Quantitative data were analyzed using descriptive statistics (mean, standard deviation) and a paired-samples t-test to determine whether significant differences existed between the bilingual and French-only test scores.

Qualitative data from the focus group were analyzed thematically following Braun and Clarke's (2006) framework, involving familiarization, coding, and identification of recurrent patterns

The qualitative insights will be used to triangulate the quantitative findings and provide a deeper understanding of how language proficiency influences diagnostic assessment results.

2.6 Ethical Considerations

Ethical approval was obtained from the school administration. Students were informed that their participation was voluntary, that their scores would not affect their grades, and that data would remain anonymous. Confidentiality and ethical integrity were maintained throughout the study.

3. Results

3.1 Quantitative Findings

The analysis compared students' performance on the two diagnostic tests to determine the extent to which the language of assessment affected their results.

Table 1: *Table of Descriptive Results across Both Tests*

Test	Mean	SD	Min	Max
Arabic French version score 1	*9,91*	*3,89*	*0*	*16*
French version score 2	5,20	3,25	0	11

Descriptive statistics revealed that mean scores were higher on the bilingual (French–Arabic) M=9,91 test than on the monolingual French version M=5,20. While students demonstrated an overall medium understanding of the scientific concepts in both tests, their performance declined when the Arabic translations were removed.
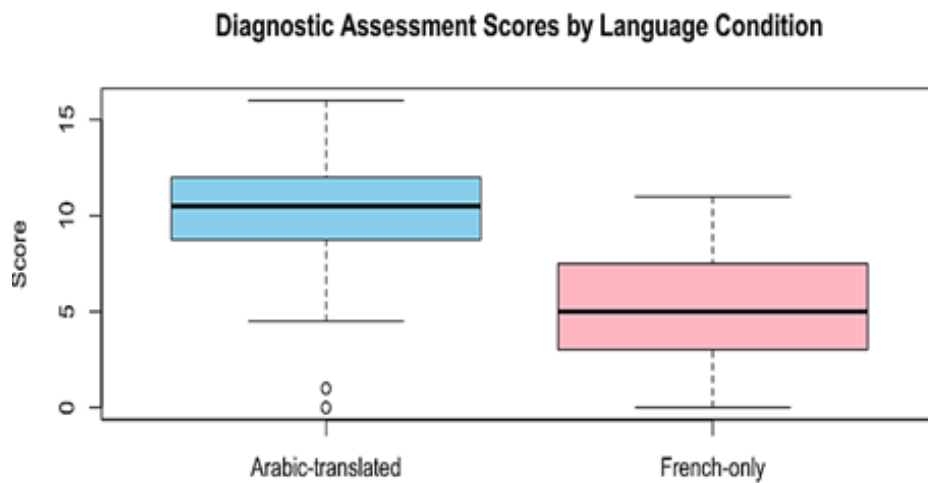
Figure 1 *Boxplot of Diagnostic Assessment Scores by Language Condition*

A paired-samples t-test confirmed that the difference between the two sets of scores was statistically significant ($t = 5.40$, $p < .001$), indicating that the language of assessment had a measurable impact on students' diagnostic performance. As shown in Figure 1, the median score in the bilingual version was higher than in French-only version, with fewer low outliers, reinforcing the significant mean difference found in the paired t-test ($p < .001$).
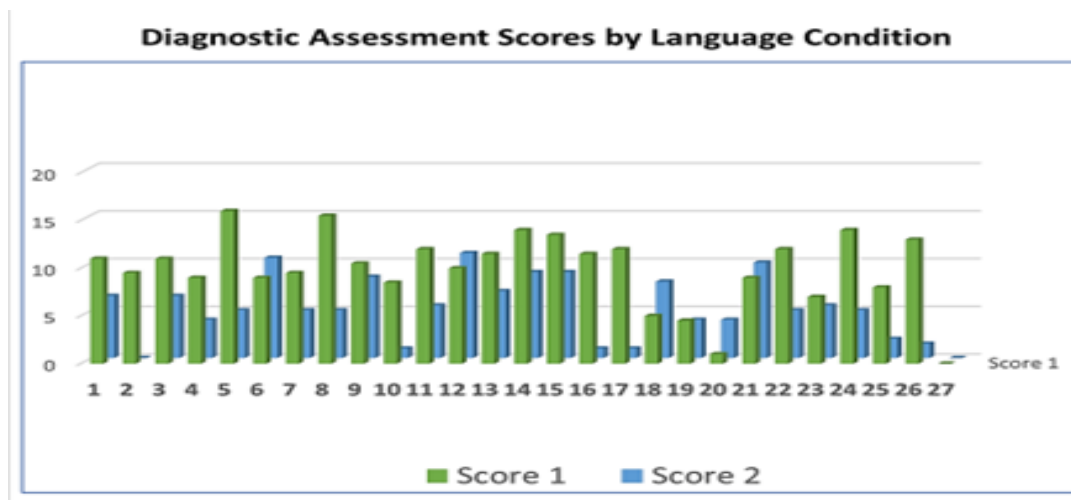

Figure 2. *Individual Diagnostic Assessment Scores in Bilingual (Arabic–French) and French-only Conditions*

Most students scored higher in the bilingual version, confirming the significant mean difference found in the paired t-test (p < .001).

This suggests that comprehension difficulties related to the language of the test rather than content mastery likely influenced outcomes although the general performance was medium in both tests. Adding Arabic translation to the diagnostic test improved student's performance. However, the correlation between both versions was weak (r = .21) and non-significant, indicating that performance in the two language conditions was not strongly aligned.

3.2 Qualitative Findings

The focus group discussion provided additional context for interpreting these results. Thematic analysis demonstrated three recurring themes illustrating how linguistic factors influenced student performance:

3.2.1 Comprehension Difficulties

Most students reported understanding the general scientific concepts but struggling to interpret complex or technical French phrasing in the test. They described the bilingual version as "easier to follow" and "less stressful." Several students noted that they could "recognize the questions faster" when Arabic terms were included. This difficulty was particularly evident in higher order cognitive demand tasks such as suggest, describe, or explain as they required more elaborated responses. In contrast, students reported fewer problems with objective items such as QCM and ordering.

3.2.2 Vocabulary Barriers

Students highlighted challenges in recalling or translating specific scientific terms from Arabic to French during the monolingual test. One participant explained that "we know the answer in Arabic, but the French words confuse us." These vocabulary gaps often led to hesitation, misinterpretation of the questions instructions, or incomplete responses, affecting their scores in the second test (French version).

3.2.3 Recognition of Test Similarity

Interestingly, only a few students realized that both tests assessed the same content. Most perceived the second (French-only) test as "different and more difficult," suggesting that language influenced their perception of task difficulty, even when the underlying scientific content was unchanged. The data supports the idea that linguistic barriers significantly affect diagnostic assessment performance in science classes.

Quantitative evidence demonstrated lower scores in the French-only condition, while qualitative insights explained why: students' comprehension and vocabulary limitations interfered with their ability to demonstrate understanding. This alignment between trends and students' experiences reinforces the validity of the interpretation that language proficiency not conceptual weakness is a key factor shaping diagnostic outcomes. Hence affecting teacher assessment judgment.

4. Discussion

The findings of this study highlight the important role of language in shaping students' diagnostic assessment performance in science classes. Quantitative analysis revealed a significant improvement when Arabic translations accompanied French test items ($t = 5.40$, p $<$ .001), suggesting that comprehension rather than content mastery was the main barrier faced by students and influencing diagnostic assessment outcomes.

Thematic insights from students further reinforced this interpretation, indicating that difficulties primarily from linguistic complexity, and anxiety related to French only tasks. Together, these results emphasize that diagnostic validity depends not only on the quality of the assessment instrument designed but also on its linguistic accessibility within multilingual educational contexts such as Morocco.

Our findings align with those of other authors. Tarnaoui (2018) states that many students fail not because of the scientific content itself, but due to difficulties in understanding the language of instruction a conclusion that perfectly corresponds with our results. In the same line, Haidar (2024) observes that students often struggle to grasp scientific concepts in French because they do not master the language itself and face a technically complex vocabulary specific to scientific discourse. This idea is also echoed by Azennoud et al. (2025), who emphasize that students in science classes, in particular, have a hard time with French, which makes it much harder for them to understand, participate, and do well overall.

Students noted that the French version is easier to follow and less stressful, this finding is echoed in an early work of MacIntyre, P. D., & Gardner, R. C. (1989) as they found that "communicative anxiety" significantly influence performance indicating a complex and unclear link between anxiety and language proficiency in both the learning and the production of French vocabulary. Similarly, a recent study conducted in Algeria Al-Qadri et al., (2023) found that

language proficiency and multilingual background significantly influenced students' anxiety levels. Students with stronger foreign language skills had lower anxiety, however those with limited proficiency particularly in the language of instruction reported even a higher stress during learning and assessment. In an Interesting way, the multilingual students tended to feel less anxious, suggesting according to the authors familiarity with languages may ease the cognitive and emotional load associated with studying in a non-native language.

Taken together, these results underline that linguistic factors and emotional responses are intertwined in shaping diagnostic assessment outcomes. For multilingual contexts such as Morocco, ensuring that assessment tools are both linguistically accessible and affectively supportive is essential for promoting fairness and accurately identifying students' true scientific understanding.

5. Pedagogical Implications

These findings remind us that teaching and assessment are never 100 % neutral acts; they are shaped by various factors such as the language through which knowledge is transmitted and assed. For science teachers in Morocco, this means that diagnostic assessments must account for students' linguistic realities as much as their scientific understanding of concepts. Integrating brief clarifications, or translated keywords for complicated scientific terms or instructions, can help students demonstrate truly what they actually know rather than what they can translate successfully to French. Teacher training should also include explicit attention to the role of language in assessment, especially in contexts where language remains a challenge for many learners. Above all, the results suggest that improving test validity in multilingual classrooms does not require lowering cognitive demand, but ensuring that students are assessed on ideas not on their fluency in a language.

6. Conclusion

This study set out to explore how the language of assessment shapes students' diagnostic performance in Moroccan science classrooms. The results revealed that the difference in scores between the bilingual and French-only tests was not a matter of knowledge, but of language. Students understood the scientific content but were held back by linguistic and emotional barriers. These findings reinforce the idea that fair assessment begins with linguistic accessibility. In

multilingual contexts like Morocco, recognizing the role of language in evaluation is not simply a pedagogical detail it is a condition for equity and genuine learning and a true assessment judgement.

7. Limitations

This study has several limitations. The sample size was relatively small and drawn from a single school, which limits the generalizability of the findings. Moreover, although care was taken to keep both test versions equivalent, translation effects may have slightly altered item difficulty making it more appealing. Finally, the qualitative data did rely on students' self-reported perceptions, which may not fully the exact capture of the complexity of their cognitive or emotional experiences during testing. Despite these limits, the findings provide meaningful insight into how language can influence assessment validity and student performance in multilingual contexts.

References

Al-Qadri, A. H., Boukheir, M., & Belmokhtar, S. (2023). Language anxiety in an Algerian multilingual tertiary context. Humanities and Social Sciences Communications, 10(1), 109.

https://doi.org/10.1057/s41599-023-01594-1

Azennoud, A., Lamghari, K., & Guaad, A. (2025). The impact of French as a medium of instruction on academic access and performance in Moroccan higher education. Moroccan Journal of Quantitative and Qualitative Research, 7(3).

https://doi.org/10.48379/imist.prsm/mjqr-v7i2.54966

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139–148.

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

Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique. (2018). L'enseignement supérieur au Maroc : efficacité, efficience et défis du système universitaire à accès ouvert (Rapport sectoriel). Rabat: CSEFRS.

Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique (CSEFRS). (2019). Mise en œuvre de la Loi-Cadre 51-17 sur l'éducation, la formation et la recherche scientifique. Rabat: CSEFRS.

El Amrani, A. (2013). Langue et société : Arabisation et fracture linguistique au Maroc. Casablanca: Éditions Le Fennec.

Elmadhi, A., & Boukhch, M. (2025). L'enseignement bilingue de l'éveil scientifique au cycle primaire en contexte éducatif marocain: Analyse de contenu des textes officiels. European Scientific Journal, 21(5), 133.

https://doi.org/10.19044/esj.2024.v21n5p133

Fadlillah, N., & Kusaeri, K. (2024). Optimizing assessment for learning in Islamic education through authentic and diagnostic assessment: A systematic literature review. Jurnal Kependidikan, 10(2), 654–667.

https://doi.org/10.33394/jk.v10i2.11555

Haidar, M. (2024). L'enseignement scientifique en français au Maroc : défis linguistiques et pédagogiques des lycéens marocains. Revue des Sciences de l'Éducation et de la Formation, 9(2), 77–95.*

Ismaili, O., et al. (2024). Le français dans les classes BIOF au Maroc : une discipline ou des disciplines ? Revue Internationale du Chercheur, 5(3), 562–583.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. Language Learning, 39(2), 251–275. https://doi.org/10.1111/j.1467-1770.1989.tb00423.x

OECD. (2005). Formative assessment: Improving learning in secondary classrooms. Paris: OECD Publishing.

OECD. (2019). PISA 2018 results (Volume I): What students know and can do — Country note: Morocco. Paris: OECD Publishing. https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/pisa/publications/national-reports/pisa-2018/featured-country-specifi overviews/PISA2018_CN_MAR.pdf

OECD. (2023). PISA 2022 results (Volume I & II): Country note: Morocco. Paris: OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/11/pisa-2022-results-volume-i-and-ii-country-notes_2fca04b9/morocco_4c81477c/10dfcb74-en.pdf

OECD/UNESCO. (2021). Éducation au Maroc : Analyse du secteur. Paris: OECD Publishing.

Scriven, M. (1983). Evaluation thesaurus. Sage Publications.

Scriven, M. (1991). Evaluation thesaurus (4th ed.). Sage Publications. https://doi.org/10.4135/9781412986420

Tang, F., & Zhan, P. (2021). Does diagnostic feedback promote learning? Evidence from a longitudinal cognitive diagnostic assessment. AERA Open, 7(1). https://doi.org/10.1177/23328584211060804

Tarnaoui, L. (2018). L'enseignement des sciences en français au Maroc : entre réformes linguistiques et réalités pédagogiques. Revue Marocaine de l'Éducation et de la Formation, 4(1), 44–59.*

UNESCO. (2015). Education for all 2000–2015: Achievements and challenges. Paris: UNESCO.