

Designing and Implementing Automated Systems for the Evaluation of Linguistic Competences

Bedraoui Noura ¹

Faculty of Letters and Foreign Languages Amar Telidji University of Laghouat, Algeria

Email address: noura.badraoui.fr@lagh.univ.dz

Received : 21/07/2025 ; Accepted : 25/12/2025 ; Published : 11/02/2026

Abstract

The integration of artificial intelligence into university, level foreign language education opens up a novel and highly consequential field of inquiry centered on the assessment of linguistic competences, a domain historically grounded in the teacher's expert judgment and in evaluative frameworks shaped by normative and standardized practices. Within the context of ongoing techno, pedagogical transformation, the design and implementation of automated assessment systems emerge as powerful levers for reshaping evaluative practices, while simultaneously raising major theoretical, methodological, and ethical questions.

This paper aims to examine the principles, modalities, and effects of automated systems for assessing linguistic competences in higher education, drawing on advances in natural language processing and machine learning algorithms. It seeks to demonstrate that these systems, capable of analyzing large, scale corpora of written and oral productions, enable a more nuanced and continuous evaluation of linguistic competences based on morphosyntactic, lexical, discursive, and pragmatic indicators. In this respect, assessment automation offers expanded possibilities for immediate feedback, individualized learning monitoring, and the personalization of educational pathways, thereby addressing contemporary challenges related to massification and equity in higher education.

However, the algorithmic rationalization of assessment cannot be conceived as a neutral or transparent process. Adopting a critical stance, this study interrogates the epistemological presuppositions embedded within automated assessment systems by questioning the implicit conceptions of linguistic competence they encode. It examines which dimensions of language are foregrounded and measured, and which are, conversely, marginalized or rendered invisible. This analysis highlights the tensions between the promise of algorithmic objectivity and the intrinsic complexity of language practices, which are shaped by variation, creativity, and socio, cultural embeddedness.

Furthermore, the study explores the reconfiguration of the role of the language teacher, who is no longer positioned solely as an evaluator but increasingly as a designer of assessment dispositifs, an interpreter of AI, generated results, and a guarantor of the pedagogical meaning of evaluation. From this perspective, automated assessment is conceptualized as part of a hybrid approach that articulates the computational power of intelligent systems with the teacher's didactic expertise.

Ultimately, this paper seeks to contribute to an in, depth reflection on the conditions required for a reasoned and pedagogically meaningful use of automated assessment, understood not as a substitute for human evaluation but as a tool for renewing evaluative practices and redefining the aims of university, level language education.

Keywords: Artificial intelligence; automated assessment; linguistic competences; higher education; foreign language education.

The Integration of Artificial Intelligence in the Assessment of Linguistic Competences in University, Level Foreign Language Education

The integration of artificial intelligence into university, level foreign language education opens up a profoundly renewed field of reflection concerning the assessment of linguistic competences, a domain that has historically been grounded in the primacy of the teacher's expert judgment and in evaluative dispositifs largely shaped by standardized procedures inherited from a didactic tradition focused on the punctual measurement of learning outcomes. The emergence of intelligent technologies capable of automatically processing language production thus disrupts established assessment frameworks by introducing new logics of analysis, temporality, and decision, making, all embedded within a broader context of techno, pedagogical transformation.

Within this dynamic of change, the design and implementation of automated assessment systems emerge as structuring levers in the reconfiguration of evaluative practices in higher education. Drawing on advances in natural language processing and increasingly sophisticated machine learning algorithms, these systems enable large, scale analyses of extensive corpora of written and oral productions while mobilizing a plurality of linguistic indicators. Assessment thereby moves beyond a global or impressionistic appraisal of performance and instead relies on the fine, grained extraction of morphosyntactic, lexical, discursive, and pragmatic features, offering a multidimensional reading of learners' linguistic competences.

Such automation paves the way for continuous and dynamic forms of assessment capable of transcending the traditional model of the one, off examination. It allows for the generation of immediate feedback, the construction of individualized linguistic profiles, and the longitudinal monitoring of learning trajectories, thereby contributing to an increased personalization of educational pathways. In this respect, automated assessment systems appear to respond to contemporary challenges associated with the massification of higher education, while simultaneously promoting principles of equity and traceability in the evaluation of linguistic performance.

Nevertheless, this algorithmic rationalization of assessment cannot be apprehended as a neutral, transparent, or purely technical process. The present communication adopts a resolutely critical stance aimed at interrogating the epistemological presuppositions underpinning such dispositifs. Any computational modeling of language necessarily rests on implicit theoretical choices: certain dimensions of linguistic competence are foregrounded and measured with precision, whereas others, such as linguistic creativity, stylistic singularity, or the socio, cultural embeddedness of discourse, risk being partially rendered invisible. This tension highlights the persistent gap between the promise of algorithmic objectivity and the intrinsic complexity of language practices, which resist full reduction to sets of entirely quantifiable variables.

Moreover, the introduction of automated assessment contributes to a substantial redefinition of the role of the university language teacher. No longer confined to the function of evaluator, the teacher is progressively repositioned as a designer of hybrid assessment dispositifs, a critical interpreter of outputs generated by intelligent systems, and a guarantor of the pedagogical meaning of the evaluative act. The teacher thus becomes a mediating instance, responsible for articulating the computational power of artificial intelligence with didactic expertise grounded in a nuanced understanding of learning processes and instructional contexts.

Ultimately, this study seeks to contribute to an in, depth reflection on the conditions required for a reasoned, ethical, and pedagogically fruitful use of automated assessment in university, level foreign language education. Far from being conceived as a substitute for human evaluation, artificial intelligence is here envisaged as an instrument for transforming and renewing evaluative practices, one capable of participating in a redefinition of the very aims of language education in higher education, through an ongoing dialogue between technological innovation and humanistic exigency.

Research Problem

The irruption of artificial intelligence into the field of university, level foreign language education, and more specifically into the assessment of linguistic competences, constitutes an epistemological rupture that calls for a reconsideration of the very foundations of the evaluative act. Long conceived as an interpretive practice grounded in the teacher's expert discernment, assessment is now increasingly mediated by algorithmic dispositifs claiming objectivity, reliability, and the standardization of judgment.

This shift raises, with particular acuity, the question of the extent to which automated assessment, by modeling language through computational parameters, can adequately account for the complexity of linguistic competences, which are intrinsically marked by variation, creativity, and the socio, cultural anchoring of language use.

Accordingly, the central research question of this study may be formulated as follows: **how does the integration of artificial intelligence into the assessment of linguistic competences at the university level reconfigure the theoretical, methodological, and axiological frameworks of evaluation, and to what extent can automated systems reconcile the demand for algorithmic rationalization with the interpretive and humanistic dimension inherent in language education?** This interrogation entails both a critical and a prospective reflection on the conditions under which technologically assisted assessment might be conceived not as a mere operation of measurement, but as part of an expanded and contextualized understanding of linguistic competence.

Research Hypotheses

On the basis of this problematization, several structuring hypotheses guide the analytical approach of the present study.

The first hypothesis posits that automated assessment of linguistic competences, as currently designed through tools from natural language processing and machine learning, tends to privilege formalizable dimensions of language, namely morphosyntactic, lexical, and, to a certain extent, discursive features, at the expense of aspects that are more resistant to computational modeling, such as creativity, stylistic variation, and the pragmatic and cultural load of language production. This implicit selection of evaluable objects would thus contribute to a partial, if not reductive, redefinition of the very notion of linguistic competence.

The second hypothesis argues that, far from guaranteeing absolute objectivity, automated assessment systems embed theoretical and ideological presuppositions derived from the linguistic and pedagogical models that underpin their design. Algorithmic objectivity would therefore stem less from intrinsic neutrality than from a rationalization of interpretive choices, which remain contingent upon training corpora, weighting criteria, and the educational purposes assigned to these dispositifs.

A third hypothesis conceives the introduction of artificial intelligence into assessment not as a substitution for the teacher's judgment, but as a factor in the reconfiguration of the teacher's role. The language teacher is thus called upon to act as a critical mediator between algorithmically generated outputs and the pedagogical realities of the classroom, ensuring coherence between quantitative data produced by intelligent systems and the qualitative interpretation indispensable to the support of learning processes.

Finally, the last hypothesis maintains that only a hybrid approach, one that articulates the analytical power of artificial intelligence with the didactic and ethical expertise of the teacher, is capable of grounding an assessment that is genuinely formative, equitable, and context, sensitive. From this perspective, automated assessment does not constitute an end in itself, but rather an instrument in the service of a redefinition of the aims of university, level language education, oriented toward the development of a plural, reflective, and socially situated linguistic competence.

Articulated Theoretical Framework

The analysis of automated assessment of linguistic competences in university, level foreign language education is situated at the intersection of several theoretical fields whose articulation allows for a nuanced apprehension of the complexity of the phenomenon under study. First, this research draws on insights from language didactics, which conceives linguistic competence not as a mere aggregation of formal knowledge, but as a dynamic capacity to mobilize linguistic, discursive, and pragmatic resources in communicative situations that are socially and culturally situated. Rooted in the communicative and action, oriented paradigms, this approach foregrounds the contextualized, evolving, and interpretive nature of language performance.

Overlaying this didactic perspective is a second theoretical axis stemming from the language sciences, and more specifically from research on linguistic modeling and natural language processing. Automated assessment systems rely on models that operate through abstraction and formalization of language use, selecting linguistic features deemed relevant and measurable. This necessary reduction inherent to any computational modeling entails implicit theoretical choices regarding the definition of language, normativity, and competence. Far from being merely technical tools, these systems thus constitute theoretical objects in their own right, carrying a particular vision of language and its uses.

A third axis of reflection is grounded in a critical approach to educational technologies, attentive to the epistemological and axiological stakes of assessment automation. This perspective invites scrutiny of the claim to objectivity often associated with algorithms, by highlighting the constructed and situated character of computational decisions. Automated assessment then emerges as a site of tension between technical rationalization and pedagogical interpretation, a tension that calls for reflection on the ethical and didactic conditions governing its integration into university practices.

Finally, the articulation of these frameworks makes it possible to conceive automated assessment within a hybrid logic, in which artificial intelligence does not replace human judgment but reconfigures its modalities. Within this configuration, the language teacher occupies a central position as a mediating instance, capable of endowing algorithmic

outputs with pedagogical meaning and embedding them within a formative approach oriented toward the development of learners' competences.

Methodological Section

Building on the foregoing theoretical articulation, the methodological architecture of this study adopts a qualitative and analytical orientation, enriched, where analytically meaningful, by selected quantitative elements, so as to grasp both the internal functioning of automated assessment dispositifs and their effects on evaluative practices within university, level foreign language education. Such a stance is warranted by the very nature of the object under scrutiny, located at the interface of technology, language, and pedagogy, and therefore resistant to any single, method reduction.

In a first phase, the inquiry proceeds through a conceptual and critical analysis of automated assessment models grounded in natural language processing and machine learning. The aim is to examine the operative principles governing these systems, the types of linguistic data they mobilize, and the indicators they privilege in measuring language competence. This analytical work is designed to bring to the surface the implicit theoretical presuppositions embedded in the dispositifs themselves, and to identify, with precision, which dimensions of language they render legible, and which, conversely, they tend to marginalize or leave unaccounted for.

In a second phase, the study draws on a corpus of university students' linguistic productions, both written and oral, collected from evaluative situations in foreign language learning contexts. These productions are treated as situated discursive objects whose meaning cannot be severed from the pedagogical conditions of their emergence. They constitute, accordingly, a privileged empirical ground for confronting algorithmically generated outputs with a didactic and interpretive reading carried by the teacher. The objective of this confrontation is not to stage an antagonistic comparison between "machine" and "human" judgment, but rather to map convergences and divergences between these two regimes of appraisal, especially with regard to discursive coherence, pragmatic adequacy, and linguistic creativity, which often exceed strictly formalizable criteria.

Furthermore, particular attention is devoted to the analysis of the teacher's role within these hybrid dispositifs. Through the examination of evaluative practices and modes of appropriation of digital tools, the study investigates the ways in which teachers interpret, adjust, corroborate, or contest algorithmic results. This dimension makes it possible to apprehend the pedagogical mediation processes at work and to assess the extent to which assessment automation contributes to transforming professional postures, responsibilities, and decision, making economies in higher education.

Finally, all the data are situated within a reflexive perspective aimed at delineating the conditions for a reasoned and pedagogically meaningful use of automated assessment. The objective is not to measure the technical performance of artificial intelligence systems per se, but rather to interrogate their didactic relevance and their compatibility with an expanded, contextualized, and humanistic conception of language education at the university level.

Research Objectives

Within the horizon thus delineated, the present work pursues several complementary objectives.

First, it seeks to analyze the theoretical foundations and functional logics of automated assessment dispositifs designed to evaluate linguistic competences in university, level foreign language education. More specifically, the study aims to elucidate how these systems model language competence, which linguistic parameters they prioritize, and through which procedures they generate evaluative judgments.

Second, the research examines the pedagogical and didactic effects of integrating such dispositifs into university assessment practices. In doing so, it investigates the extent to which automated assessment contributes to reshaping feedback modalities, learning monitoring practices, and the personalization of training pathways, particularly in contexts marked by large, scale student cohorts and institutional demands for efficiency.

Third, a transversal objective consists in critically interrogating the epistemological and ethical implications of assessment automation, by foregrounding the tensions between algorithmic rationalization and human interpretation, as well as the conditions under which a hybrid and reasoned use of artificial intelligence may be instituted in the service of language education.

Corpus of the Study

In accordance with these objectives, the corpus selected for the present study consists of linguistic productions by university students enrolled in foreign language programs. It includes, on the one hand, written productions stemming from formal academic evaluative situations, argumentative essays, syntheses, extended responses, and, on the other hand, oral productions collected in contexts of evaluative interaction, presentations, spontaneous speech, and dialogic exchanges.

These data are approached as contextualized language objects embedded within specific pedagogical practices. The corpus is constructed so as to reflect a diversity of proficiency levels, discursive genres, and communicative situations, thereby enabling a nuanced analysis of linguistic performances and of the ways in which automated assessment systems process them.

Methodological and Analytical Tools

The analysis mobilizes a complementary set of tools, articulating artificial intelligence technologies with didactic instruments of qualitative interpretation. Automated assessment dispositifs rely on natural language processing modules designed to extract and analyze linguistic features at multiple levels, morphosyntactic, lexical, discursive, and pragmatic. These tools are used to generate scores, linguistic profiles, and progression indicators.

In parallel, qualitative analytical grids grounded in language didactics frameworks are employed to interpret students' productions. These grids make it possible to contextualize algorithmic outputs and to examine their pedagogical relevance by taking into account communicative intentions, discursive coherence, and socio, pragmatic adequacy. The articulation of these two categories of instruments aims to establish an analytical dialogue between automated measurement and expert interpretation, an essential condition for a hybrid and formative assessment model.

Criteria for Analysis and Assessment

The criteria retained for the analysis of linguistic productions are organized according to a multidimensional conception of language competence. At the formal level, assessment focuses on morphosyntactic accuracy, lexical richness and precision, as well as sentence, level and textual structuring. These dimensions correspond to the aspects most readily formalizable and thus most easily handled by automated systems.

At the discursive and pragmatic level, the analysis attends to the global coherence of productions, argumentative organization, genre adequacy, and sensitivity to the communicative context. These dimensions, more deeply dependent on situated interpretation, constitute a privileged space for interrogating the limits of automated assessment.

Finally, a set of transversal criteria is mobilized to examine the pedagogical effects of automated assessment, notably in terms of feedback quality, the intelligibility of results for learners, and the contribution of automated outputs to individualized learning monitoring. These criteria aim to evaluate not only the technical validity of the dispositifs, but also their didactic relevance and their compatibility with a formative and humanistic conception of evaluation.

Reflective Methodological Conclusion

At the end of this methodological itinerary, it becomes clear that analyzing automated assessment of linguistic competences cannot be reduced either to a descriptive inventory of technical tools or to a mechanical comparison between human judgment and algorithmic output. The approach adopted here, grounded in the articulation between computational instruments and expert didactic reading, has sought, precisely, to apprehend assessment as a space of mediation, where heterogeneous logics, technical, pedagogical, and epistemological, encounter one another, and where their interaction conditions both the scope and the limits of artificial intelligence in university, level foreign language education.

The elaboration of an operational schema integrating objectives, corpus, tools, and criteria has made it possible to delineate, with rigor, an analytical framework capable of capturing the plurality of dimensions constitutive of linguistic competence. By privileging a multidimensional approach, the methodology has enabled a joint examination of formalizable aspects of language, readily processed by automated systems, alongside interpretive, discursive, and pragmatic dimensions that continue to depend on the teacher's pedagogical discernment. Rather than implying an implicit hierarchy between human and algorithmic assessment, this choice has aimed to illuminate zones of convergence as well as points of friction between these two regimes of appraisal.

Moreover, the systematic confrontation between outputs produced by automated dispositifs and qualitative analysis of student productions has shifted the analytical focus from technical efficiency to didactic pertinence. In this respect, the methodology does not pursue an instrumental validation of artificial intelligence tools; instead, it adopts a critical perspective attentive to meaning effects, normative implicatures, and professional reconfigurations induced by their integration into university assessment practices.

Such a reflexive posture constitutes a necessary precondition for the analysis of results. It invites us to read data not as objective and self-sufficient indicators, but as situated productions stemming from theoretical, methodological, and technological choices that must themselves be interrogated. The analysis that follows will therefore examine, on the basis of the constituted corpus, the manner in which automated assessment dispositifs render certain dimensions of linguistic competence visible while leaving others in the shadow, and it will evaluate the pedagogical significance of these gaps.

It is within this perspective that the analysis of results is opened, conceived not as a mere restitution of figures or scores, but as a space of critical interpretation aimed at understanding how artificial intelligence, as an evaluative tool, participates in redefining practices, roles, and aims in university-level foreign language education.

Opening: Future Research Perspectives

By way of an opening, this reflection invites several research avenues capable of extending and deepening the analysis. First, a prospective axis would interrogate the capacity of AI systems to integrate dimensions that currently remain marginal within automated assessment, such as discursive creativity, stylistic variation, or the subjective appropriation of linguistic norms. This orientation would require rethinking learning models not solely in terms of conformity, but by taking into account the plurality of uses and the meaningful deviations that characterize university language practices.

A second avenue would examine the articulation between automated assessment and formative evaluation through longitudinal dispositifs. Over the duration of a university curriculum, it would be necessary to investigate how AI-generated feedback influences learning trajectories, students' linguistic strategies, and their relationship to error, thus shifting attention from punctual performance to the processes of linguistic appropriation.

Equally essential is the study of teachers' practices. The integration of AI into assessment entails a transformation of professional competences that remains insufficiently explored empirically. Analyzing how teachers appropriate these tools, adapt them to pedagogical aims, or resist certain aspects would clarify the institutional and didactic conditions required for meaningful deployment.

Finally, a broader reflection could address the ethical and political stakes of assessment automation in higher education. Questions of algorithmic transparency, protection of students' linguistic data, and fairness of evaluative decisions deserve to be considered not merely as regulatory constraints, but as research objects in their own right, at the intersection of language sciences, pedagogy, and the digital humanities.

Thus, rather than closing the debate, this communication seeks to open a durable space of inquiry in which artificial intelligence, rather than imposing itself as a technological norm, may become a terrain for critical experimentation in the service of language education attentive to the complexity of language and to the singularity of learning trajectories.

References (APA Style – 7th Edition)

1. Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
2. Chapelle, C. A., & Voss, E. (2016). Evaluation of language learning technologies. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 378–392). Routledge. <https://doi.org/10.4324/9781315657899>
3. Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
4. De Jong, N. H. (2018). Fluency in second language assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 1–16). Springer. https://doi.org/10.1007/978_3_319_02261_1_25

5. Davis, N. L., Gough, M., & Taylor, L. L. (2019). Online teaching: Advantages, obstacles and tools for getting it right. *Journal of Teaching in Travel & Tourism*, 19(3), 256–263. <https://doi.org/10.1080/15313220.2019.1612313>
6. Fulcher, G. (2010). *Practical language testing*. Hodder Education.
7. Khalil, M., & Ebner, M. (2017). Learning analytics: Principles and constraints. In D. Lang et al. (Eds.), *Learning analytics in higher education* (pp. 178–190). Springer. https://doi.org/10.1007/978_3_319_52977_6_14
8. Knoch, U. (2018). Automated writing evaluation. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (pp. 1–16). Springer. https://doi.org/10.1007/978_3_319_02261_1_20
9. McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.
10. Perelman, C., & Olbrechts, Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation* (J. Wilkinson & P. Weaver, Trans.). University of Notre Dame Press. (Original work published 1958)
11. Reeves, T. C., & Reeves, P. M. (2015). Educational technology research in a VUCA world. *Educational Technology*, 55(2), 26–30.
12. Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Longman.
13. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
14. Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: Digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, 45(2), 107–114. <https://doi.org/10.1080/17439884.2020.1761641>
15. Zhai, X., He, P., & Xu, J. (2021). Artificial intelligence in education: A critical review. *IEEE Transactions on Learning Technologies*, 14(4), 481–497. <https://doi.org/10.1109/TLT.2021.3054478>