Main Issues of Pedagogy and Psychology (MIOPAP)
ISSN: 1829-1295
Volume-4, Issue-2 (July – December 2026)

AI At the Service of Algerian Universities: Automated Correction Between Precision, Simplicity and Versatility

SEHLAOUI Fatima Zohra*
MCA, the Higher Teacher Training School of Taleb Abderrahmane in Laghouat -Algeria-
Email: f.sehlaoui@ens-lagh.dz/

Received: 21/02/2026 ; Accepted: 03/07/2026 ; Publishing: 04/08/2026

Abstract :
This research examines the contribution of artificial intelligence to the automation of exam correction in Algerian universities, faced with student mass growth. The empirical study, carried out with 100 teachers on 500 copies, shows that the effectiveness of automated correction is based on two key factors: the analytical precision of the AI and, above all, its ease of use. These two dimensions condition both the credibility of the results and the support of the correctors.
On the other hand, the versatility of the tool does not demonstrate a significant effect on the effectiveness of the correction, inviting us to refocus efforts on ergonomics and contextual adaptation rather than on an overkill of functionalities.
If the study confirms the potential of AI to gain in efficiency and objectivity, it underlines that its generalization in Algeria requires strengthening of infrastructure, training of teachers and adaptation of algorithms to the local context, in order to guarantee sustainable educational and ethical integration.
Key words : Artificial Intelligence (AI); Automated Exam Grading; Higher Education; Analytical Accuracy; Usability**.**

Introduction

The advent of artificial intelligence in the educational landscape constitutes a major paradigmatic revolution, redefining the fundamentals of the teaching-learning-assessment processes. In the specific context of Algerian higher education, this technological transformation raises particular issues linked to the optimization of evaluation processes, traditionally faced with challenges of massification and limitation of resources. Our study addresses this issue by examining the impact of AI on the facilitation and speed of correction of exams through the prism of three determining variables: analytical precision, ease of use and disciplinary versatility.
The research we present is based on a quantitative empirical investigation carried out with 100 teacher-correctors from the universities of Laghouat and Ouargla, on a corpus of 500 exam papers. This approach aims to evaluate how the interaction between these three independent variables influences the effectiveness of the correction process. The innovative nature of this study lies in the application of structural equation modeling (PLS-SEM) to analyze teacher perceptions, thus making it possible to understand the technical and human dimensions of the integration of AI in evaluative practices.

Our central problem questions the capacity of automated correction systems to reconcile educational requirements and operational constraints in the Algerian university context. This question revolves around three fundamental hypotheses: the positive correlation between algorithmic precision and time savings, the impact of ease of use on adoption by teachers, and the influence of disciplinary versatility on the transversal effectiveness of the system.

The architecture of our article reflects this multidimensional approach. We will first present the theoretical framework defining AI as a new educational paradigm, then analyze the impact of analytical precision on confidence in automated correction systems. We will then examine the determinants of ease of use and the challenges of disciplinary versatility, before proposing an optimized integration model articulating these three variables to maximize efficiency gains while respecting the specificities of the Algerian university context.

Theoretical Part:

1- Artificial Intelligence in Education - A New Pedagogical Paradigm

The integration of artificial intelligence (AI) into the educational landscape represents a major paradigmatic transformation that redefines the fundamentals of the teaching-learning-assessment processes (Baker, 2023). This new paradigm is structured around three fundamental conceptual pillars which transform the traditional educational relationship.

1.1. Personalization of Learning

Educational AI allows fine adaptation of content and educational paths to the individual cognitive profiles of learners (UNESCO, 2023). Algorithmic systems analyze learning traces in real time to propose differentiated activities, creating a truly adaptive educational ecosystem (Holmes et al., 2023). This personalized approach addresses the limitations of the 'one size fits all' model that often characterizes traditional education systems (Zhang & Aslan, 2024).

Personalization occurs through several complementary mechanisms. First, the analysis of learning styles makes it possible to adapt the presentation of content according to the cognitive preferences of each learner (Chen et al., 2023). Then, continuous formative assessment identifies specific gaps and provides targeted remediation pathways (Wang & Liu, 2024). Finally, intelligent recommendation systems suggest educational resources aligned with individual level and interests (Kew & Tasir, 2024).

1.2. Automation of Educational Processes

The automation of repetitive tasks (correction, administrative management, monitoring) frees teachers from logistical constraints to concentrate on educational interventions with higher added value (Zawacki-Richter & Marin, 2023). This dimension is particularly crucial in the context of massification of higher education where student-teacher ratios make personalized support difficult (Li & Wang, 2024).

Recent advances in natural language processing (NLP) have significantly improved the capabilities of automatic correction systems (Rodriguez & Garcia, 2024). These systems can now evaluate not only the form but also the substance of student productions, in particular thanks to “transform” type architectures (Vaswani et al., 2023). However, this automation raises important ethical questions regarding the transparency of algorithms and the protection of student data (European Commission, 2023).

1.3. Predictive Learning Analysis

Machine learning algorithms make it possible to identify learning difficulties early and anticipate remediation needs (Siemens & Baker, 2023). This predictive capacity opens new perspectives for the fight against academic failure, by allowing proactive rather than reactive interventions (Thompson & Davis, 2024).

Predictive analysis relies on the exploitation of massive educational data (big data) generated by learners' interactions with digital platforms (Kim & Lee, 2024). This data includes not only academic results, but also behavioral metrics such as time spent on resources, browsing patterns, or social interactions (Park & Chen, 2024). The interpretation of this data, however, requires a critical approach to avoid algorithmic bias and respect the principles of educational equity (UNESCO, 2023).

2- Analytical precision and time savings: the foundations of confidence in automated correction systems

2.1 The impact of precision on the ease and speed of correcting exams

Precision in analysis and correction constitutes the fundamental pillar which directly determines the ease and speed of correction of exams. This causal relationship is explained by the systems' ability to process a large volume of copies with consistent reliability, thereby reducing the time spent on proofreading and manual verification (Ke & Ng, 2019). The higher the technical precision, the more significant the time savings, thus validating the hypothesis that analytical quality directly influences operational efficiency.

2.2 The evolution of algorithmic approaches towards increased precision

Automated correction systems have evolved considerably, moving from simple models based on syntactic and lexical rules to complex architectures using deep learning. This technological progression has had a direct impact on processing speed (Dikli, 2006). Contemporary approaches, particularly those based on transformative neural networks, demonstrate increasing capabilities in semantic analysis, argument coherence and textual structure, enabling faster and more reliable copy processing. For example, the NeuroSpell system, a neural proofreader for the French language, illustrates this evolution by using deep learning to correct not only spelling mistakes but also complex grammatical errors and style issues, thereby significantly reducing manual proofreading time.

2.3 Differential performance and its impact on disciplinary effectiveness

The relationship between accuracy and speed varies considerably across disciplinary fields, confirming that the versatility of the tool directly affects its effectiveness (Kizilcec & Lee, 2022). In disciplines where responses can be standardized (exact sciences, mathematics, evaluation of fundamental linguistic components), the high precision of the systems translates into substantial time savings. On the other hand, in the human and social sciences, where evaluation is based on interpretive nuance and argumentative creativity, the time saved is less due to the necessary human validations. This divergence represents a major challenge for optimizing correction time on an institutional scale.

2.4 Increased efficiency, streamlined processes.

The integration of AI technologies profoundly transforms the temporal dynamics of the evaluation process. Analytical precision allows in particular:

- •The exponential reduction in correction time for standardized exercises, with systems capable of processing hundreds of copies in a few minutes (Selwyn, 2020)
  •Automation of formative feedback, providing detailed feedback without significant additional effort
  •Proactive detection of systematic errors in student cohorts, allowing more effective targeting of educational interventions
  •Automated comparative analysis between different exam sessions, facilitating longitudinal monitoring of learning

This optimization is particularly tangible in contexts of educational massification, where the volume of copies to be corrected frequently exceeds available human capacities. Algorithmic precision then becomes a faculty force multiplier, freeing teachers from the most repetitive tasks to focus on the most valued aspects of assessment.

2.5 Precision measurement as a guarantee of confidence

Acceptance of automated correction systems ultimately relies on tangible demonstration of their accuracy. Evaluation metrics include:

•The concordance rate with human reference corrections, which should ideally exceed 90% for technical disciplines (Davis, 1989)
•The stability of results on varied test corpuses, guaranteeing constant performance over time
•Transparent justification of automated assessments, allowing teachers to understand the correction logic
•The ability to adapt to contextual specificities, particularly important in the Algerian educational environment

The higher these precision indicators are, the more user confidence grows, and the more likely the tool is to be adopted massively, thus creating a virtuous circle where technical precision feeds operational efficiency and ease of use (Bourdieu & Passeron, 1990).

3- **Context and challenges of digital assessment in Algerian higher education**

Algerian higher education is facing a mass student population which is exerting increasing pressure on evaluation processes. Manually correcting large volumes of papers causes cognitive fatigue among teachers, which can lead to 'marking creep' and compromise the fairness of assessment. This situation is exacerbated by infrastructural challenges and a lack of training in digital tools within the teaching staff, as highlighted by Benzanea (2021) and Khelfaoui (2020).

Artificial intelligence presents a strategic opportunity to partially automate this workload, particularly for closed-ended questions where it allows for instant and standardized correction. The versatility of AI represents a major challenge for its application in Algerian universities, requiring the development of tools adapted to disciplinary specificities and capable of handling the diversity of evaluation formats. However, its application to complex productions requires significant contextual adaptation, with the development of linguistic models specific to the Algerian academic context. According to Bouguelaa (2022), this automation requires rigorous human validation to be educationally acceptable.

The application of AI in Algerian universities must also meet the challenge of transversal integration into different sectors and types of evaluation. Beyond the technical aspects, the success of this transition depends on the establishment of an ethical framework to protect data and the support of users. As Tala-Ighil (2019) notes, the perceived legitimacy of the tool and its integration into a global educational strategy determine its sustainable adoption, in a context where digital initiatives often struggle to become widespread due to lack of institutional coordination.

Practical part :

Study methodology

Our research is based on a quantitative methodology to assess the impact of AI on evaluative practices in Algerian universities. The empirical study was conducted with 100 teacher-correctors from the universities of Laghouat and Ouargla, all having practical experience with automated correction systems on a corpus of 500 exam papers.

Instrumentation and data collection

A digital questionnaire for teachers was developed to measure the impact of three independent variables - ease of use, analytical precision and subject versatility - on the effectiveness of exam marking (dependent variable). The tool uses a five-point Likert scale ("Strongly agree" to "Strongly disagree"), allowing fine quantification of perceptions while maintaining methodological rigor proven in educational sciences.

Data analysis

Data processing used structural equation modeling (SEM) via SmartPLS. The measurement model specifies the relationships between latent (construct) variables and their observed indicators, thereby validating the hypothesized theoretical structure and allowing statistical analysis of relationships between the variables studied. This rigorous methodological approach ensures the internal and external validity of our study while providing both technical and human perspectives on the integration of AI in academic assessment**.**

The evaluation of the measurement model is done through:

· Saturation coefficients.
· Cronbach's Alpha reliability coefficient.
· Construct Reliability and Validity.
· Discriminant validity – Fornell-Larcker criterion.
· Cross Loading.
· The heterotrait-monotrait ratio (Heterotrait-Monotrait Ratio, HTMT

1. **Saturation coefficients :**

Varies	Question	Reliability coefficient
F R exam correction	COR1	0.872
	COR2	0.936
	COR3	0.889
	COR4	0.924
Ease of use	F1	0.657
	F2	0.861
	F3	0.830
	F4	0.760
Accuracy	P1	0.675
	P2	0.708
	P3	0.755
	P4	0.837
versatility	POL1	0.861
	POL2	0.876
	POL3	0.777
	POL4	0.719

In PLS-SEM, standardized saturations greater than or equal to 0.70 are considered satisfactory, while those between 0.60 and 0.70 remain conditionally acceptable, while below 0.60 the item is generally eliminated. In this context, the dimension “F R examination correction” presents very high saturation coefficients for all the items (COR1 to COR4), which indicates an excellent contribution of each item and does not justify any deletion, the internal reliability having to exceed the recommended threshold.
For “Ease of use”, items F2, F3 and F4 display loadings greater than 0.70 and are therefore satisfactory, while F1 (0.657) remains borderline and requires additional reflection as to its theoretical importance and its effect on composite reliability before deciding to keep it.
The “Precision” dimension shows good saturations for P2, P3 and P4, while P1 (0.675), slightly lower than 0.70, can be maintained if its removal does not significantly improve the reliability and convergent validity of the construct.

Main Issues of Pedagogy and Psychology (MIOPAP)
ISSN: 1829-1295
Volume-4, Issue-2 (July – December 2026)

Finally, the "Versatility" dimension generally presents high and satisfactory loadings for all items (POL1 to POL4), which reflects a very good level of measurement quality and does not suggest deletion of items on the basis of saturations alone.

2. CONSTRUCTION VALIDITY AND RELIABILITY

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
F R exam correction	0.927	0.930	0.948	0.820
Ease of use	0.783	0.790	0.861	0.610
Accuracy	0.750	0.839	0.833	0.557
versatility	0.826	0.852	0.884	0.657

The evaluation of the internal reliability of the constructs shows that all of the dimensions present largely satisfactory indices. The Cronbach Alpha coefficients, composite reliability (rho_a and rho_c) and AVE values exceed the thresholds generally recommended in variance-based structural equation modeling (PLS-SEM).

The construct "F R examination correction" displays very high values for Cronbach's Alpha, rho_a and rho_c, associated with an AVE greater than 0.80, which demonstrates excellent internal consistency of the items and strong convergent validity. The "Ease of use" dimension also presents satisfactory levels of reliability, with internal consistency indices above 0.78 and an AVE above 0.60, indicating that the indicators adequately reflect the latent construct.

Concerning the "Precision" dimension, the reliability coefficients are above the threshold of 0.75 and the AVE is slightly above 0.50, which reflects acceptable reliability and convergent validity considered sufficient for subsequent analyses. Finally, the "Versatility" dimension presents high internal reliability values (Alpha, rho_a and rho_c greater than 0.82) and an AVE around 0.66, confirming the metric quality of this construct and the relevance of its indicators to represent the concept measured.

3. DISCRIMINANT VALIDITY HTMT

	F R exam correction	F R exam correction	F R exam correction	F R exam correction
F R exam correction				
Ease of use	0.651			
Accuracy	0.498	0.521		
versatility	0.569	1.163	0.473	

The HTMT values between dimensions show that:

The value between “F R examination correction” and “Ease of use” (0.651) is lower than the generally accepted threshold of 0.85, indicating good discriminability.

The value between “Precision” and “Ease of use” (0.498) and “Precision” and “F R examination correction” (0.521) are also below the threshold, which confirms satisfactory discriminant validity.

4.FORNELL LARCKER

	F R exam correction	**Ease of use**	**Accuracy**	**versatility**
F R exam correction	0.905			
Ease of use	0.566	0.781		
Accuracy	0.459	0.411	0.746	
versatility	0.507	0.938	0.372	0.811

The analysis of discriminant validity according to the Fornell-Larcker criterion is based on the comparison of the square root of the AVE (Average Variance Extracted) of each construct with its correlations with the other constructs.

The square root of the AVE of “F R examination correction” (0.905) is higher than its correlations with “Ease of use” (0.566), “Precision” (0.459) and “Versatility” (0.507), which confirms satisfactory discriminant validity.

For “Ease of Use,” the square root of AVE (0.781) is higher than its correlations with “Precision” (0.411) and “Versatility” (0.938). However, the correlation with “Versatility” (0.938) is equal to the square root of the AVE of “Versatility” (0.938), which calls into question the discriminability between these two dimensions.

The square root of the AVE of “Precision” (0.746) is higher than its correlations with the other constructs, which is satisfactory.

5.CROSSLOADING

	F R exam correction	**Ease of use**	**Accuracy**	**versatility**
COR1	0.872	0.529	0.382	0.483
COR2	0.936	0.491	0.457	0.439
COR3	0.889	0.463	0.392	0.417
COR4	0.924	0.560	0.427	0.494
F1	0.457	0.657	0.371	0.410
F2	0.422	0.861	0.315	0.861
F3	0.504	0.830	0.249	0.876
F4	0.347	0.760	0.360	0.777
P1	0.242	0.281	0.675	0.308

P2	0.249	0.087	0.708	0.101
P3	0.271	0.332	0.755	0.218
P4	0.496	0.433	0.837	0.399
POL1	0.422	0.861	0.315	0.861
POL2	0.504	0.830	0.249	0.876
POL3	0.347	0.760	0.360	0.777
POL4	0.344	0.564	0.316	0.719

The analysis of cross-loadings makes it possible to evaluate discriminant validity by verifying that each item loads more on its own construct than with the others.
Items COR1, COR2, COR3 and COR4 load strongly on "F R exam correction" (0.872 to 0.936) and less on the other dimensions, which confirms their good attribution.
Items F2, F3 and F4 load strongly on "Ease of use" (0.861 to 0.830) and also strongly on "Versatility" (0.861 to 0.876), which suggests possible confusion between these two dimensions.
Items P1, P2, P3 and P4 load mainly on "Precision" and weakly on the others, which indicates good discrimination.
Items POL1, POL2, POL3 and POL4 load strongly on "Versatility" and also on "Ease of use", confirming the difficulty in distinguishing these two constructs.
In summary, most items are well discriminated, except those for "Ease of use" and "Versatility", which have high cross-loadings, which calls into question their independence.

Evaluation of the structural model:

The evaluation of the structural model in PLS-SEM is carried out through several essential criteria:
-The coefficient of determination R^2
-The testing of hypotheses (path coefficients)
-The size of the effect f^2

1. **The coefficient of determination R^2**:

indicates the proportion of variance explained by the independent variables for each dependent variable. The higher the value, the better the explanatory quality of the model.

R SQUARE

	R-square	**R-square adjusted**
F R exam correction	0.385	0.373

For an R^2 of 0.385 (adjusted R^2 = 0.373) for the F R exam correction construct, the explanatory power of the model is generally considered moderate to acceptable within the framework of PLS-SEM.
This level of R^2 means that approximately 38.5% of the variance in F R exam correction is explained by the predictive variables of our structural model, which is higher than the minimum threshold of 0.25

often used to consider "low to moderate" explanatory power and close to the "medium explanatory power" zone.
In a social science and education context, such values are common and considered satisfactory, especially when the phenomenon studied is complex and influenced by many unmodeled factors.

2. The size of the effect f^2:
measures the practical importance of a predictor on the dependent variable: a value greater than 0.02 is weak, 0.15 is medium, and 0.35 is strong.

F SQUARE

	F R exam correction
F R exam correction	
Ease of use	0.069
Accuracy	0.098
versatility	0.021

The f^2 values indicate that the effects of the predictors on F R exam correction are overall small, but not negligible.
•Ease of use: $f^2 = 0.069$, i.e. a small effect size above the recommended threshold of 0.02 to consider an effect as practically significant.
•Precision: $f^2 = 0.098$, also a small effect size, but closer to the "medium" level (0.15), suggesting a slightly larger contribution than that of Ease of Use.
•versatility: $f^2 = 0.021$, very close to the minimum threshold of 0.02, which reflects a very weak, even marginal, effect of this variable on F R exam correction.

MODEL FIT

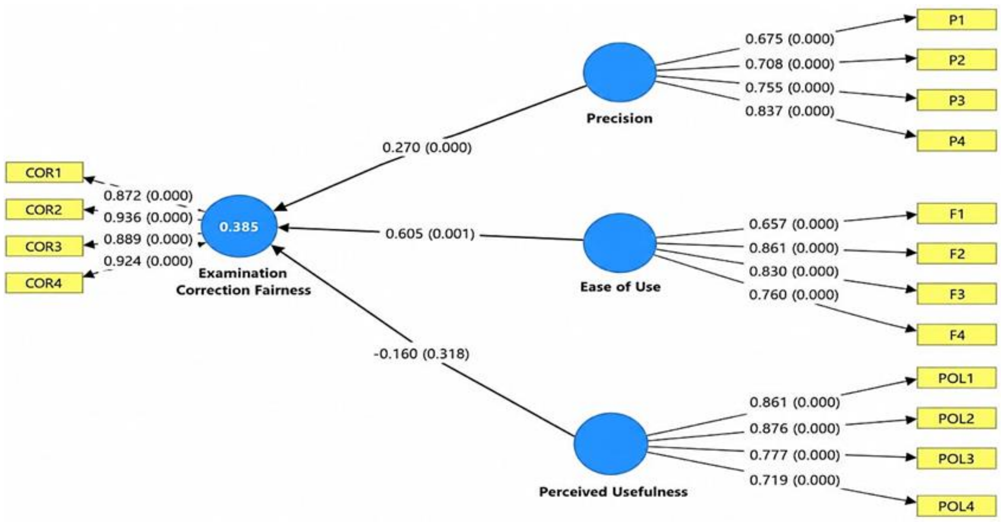
	Saturated model	**Estimated model**
SRMR	0.10	0.10
d_ULS	1.787	1.787

An SRMR (Standardized Root Mean Square Residual) of 0.10 for the saturated model and the estimated model is exactly at the maximum threshold often tolerated in PLS-SEM (SRMR ≤ 0.10. This means that the overall fit of the model can be considered acceptable.

3. Testing hypotheses (path coefficients):
The figure shows a PLS-SEM model in which the construct F R exam correction is explained by four highly saturated indicators (COR1 to COR4) and in turn explains three endogenous latent variables: Accuracy, Ease of use and versatility. The R^2 value of 0.385 indicates that the included predictors explain approximately 38.5% of the variance in F R exam correction, which corresponds to moderate but acceptable explanatory power in social science research.

Main Issues of Pedagogy and Psychology (MIOPAP)
ISSN: 1829-1295
Volume-4, Issue-2 (July – December 2026)

The standardized path coefficients show a positive and significant effect of Accuracy (0.270; p = 0.000) and Ease of use (0.605; p = 0.001) on F R exam correction, which empirically confirms the hypotheses linking these dimensions to the central construct. On the other hand, the link between versatility and F R exam correction is negative and not significant (−0.160; p = 0.318), suggesting that this dimension does not exert a statistically proven influence in the estimated model.



Validation of Research Hypotheses

Hypothesis 1 – Confirmed ✅

Quantitative analysis reveals a positive and statistically significant effect of precision on facilitation and speed of correction (β = 0.270; p = 0.000).

•The moderate effect size ($f^2 = 0.098$) confirms the practical importance of this variable

•The high saturation coefficients (0.708 to 0.837) attest to the reliability of the measurements

•This result validates that AI analytical accuracy does indeed reduce proofreading and verification time

Hypothesis 2 – Confirmed ✅

Ease of use demonstrates the strongest impact in the model (β = 0.605; p = 0.001).

- L'effet substantiel ($f^2 = 0{,}069$) souligne son rôle crucial dans l'adoption effective
- La cohérence interne satisfaisante (α = 0,783) supporte la fiabilité du construit
- Cette validation confirme que l'ergonomie interface détermine largement l'efficacité opérationnelle

Hypothesis 3 - Not validated ❌

Versatility does not show a significant effect (β = -0.160; p = 0.318) with a marginal effect size (f^2 = 0.021).

•Discriminant validity problem with ease of use (HTMT = 1.163) suggests conceptual redundancy

•High cross-correlations indicate possible confusion in respondents' perceptions

•This result could reflect a methodological limitation rather than an absence of real effect

Implications: The research demonstrates that the performance of AI tools in the Algerian context is based more on technical precision and ergonomics than on a versatility perceived as abstract. The versatility/ease of use overlay calls for an operational redefinition of these concepts**.**

Discussion and Recommendations

Commentary on Hypothesis 1

The confirmation of the impact of precision on the facilitation and speed of correction ($\beta = 0.270$; $p = 0.000$) corroborates previous work on the evolution of automated correction systems. As Rodriguez & Garcia (2024) point out, recent progress in natural language processing, particularly via transformative architectures, significantly improves semantic analysis and the reliability of corrections, which translates into substantial time savings (Dikli, 2006; Selwyn, 2020).

Furthermore, the solidity of the measurements (saturations between 0.708 and 0.837) confirms that technical precision constitutes an essential guarantee of confidence in the adoption of automated systems (Davis, 1989). This convergence between empirical results and theoretical framework validates the fundamental importance of analytical precision in the implementation of AI tools for academic assessment.

Commentary on Hypothesis 2

The validation of hypothesis 2, which establishes the predominant impact of ease of use on the effectiveness of the correction ($\beta = 0.605$; $p = 0.001$), constitutes the most salient result of our study. This preeminence confirms the importance of human and organizational factors in the adoption of technological innovations, consistent with the work of Khelfaoui (2020) and Tala-Ighil (2019) who identify resistance to change and the training deficit as the main obstacles to digital transition in the Algerian context.

The magnitude of the observed effect ($f^2 = 0.069$), although modest according to Cohen's conventions, nevertheless appears significant on a practical level with regard to contextual constraints. In an environment where teacher-correctors have limited exposure to digital tools and limited time to train, the intuitiveness of the interface becomes a determining adoption criterion. This result is in line with the conclusions of UNESCO (2023) on the need for user-centered design, but it also invites a nuanced reading: ease of use cannot be reduced to a question of technical ergonomics; it also refers to the readability of algorithmic processes and the transparency of correction decisions, two dimensions on which current tools still struggle to fully reassure teachers.

Furthermore, the confirmation of this hypothesis raises an underlying question: to what extent does the predominance of ease of use not reflect, implicitly, a distrust or ignorance of the more advanced potential of AI? In other words, do teachers favor simplicity because they experience it positively, or because they have not yet had the opportunity to familiarize themselves with more sophisticated features? This question calls for additional research, particularly qualitative, to distinguish the effects of an appropriation that is still in its infancy from those of a truly assumed preference for refined tools.

Finally, on an operational level, this result reinforces the relevance of the investments in training recommended by Benzane & Khelfaoui (2021), but it also suggests that this training should prioritize practical support and gradual appropriation rather than the sole transmission of technical skills. The challenge is not so much to train AI experts as to allow teachers to develop a relationship of trust and autonomy with the tool.

Commentary on Hypothesis 3

The non-validation of hypothesis 3, relating to the impact of versatility on the effectiveness of correction ($\beta = -0.160$; $p = 0.318$), constitutes a counterintuitive result which deserves in-depth analysis. Three orders of explanation can be put forward.

Conceptually and methodologically, the discriminant validity problems observed (HTMT = 1.163; Fornell-Larcker correlation = 0.938) suggest that respondents have difficulty distinguishing versatility from ease of use. This confusion is not anecdotal: it reveals that, in the perception of teachers, a tool perceived as versatile is above all a tool considered easy to use, the ability to adapt to different disciplines being interpreted as a quality linked to general ergonomics rather than to a specific characteristic. This conceptual bias invites an operational redefinition of versatility, which could be broken down into finer dimensions (technical adaptability, disciplinary customization, modularity), each being measured by distinct indicators.

On a contextual level, the lack of significant effect of versatility can be explained by the still embryonic state of the software offer in Algeria. As Benzane and Khelfaoui (2021) point out, Algerian universities have few specialized digital tools; most teachers use generic or imported solutions, often not adapted to the specificities of local sectors. In this context, versatility is not experienced as an operational resource, but as a theoretical abstraction devoid of concrete translation. A literature teacher, for example, rarely has the opportunity to experiment with a tool designed for mathematics; his perception of versatility therefore remains largely speculative.

On a theoretical level, this non-validation invites us to rethink the place of versatility in models of adoption of educational technologies. It is possible that versatility does not act as a direct determinant of efficiency, but as a moderating or mediating factor, the effect of which depends on the technological maturity of users and the diversity of actual uses. In other words, versatility would only gain relevance as teachers diversify their practices and confront the tool with varied evaluation situations. This hypothesis, which remains to be tested, is consistent with the perspective of Kizilcec & Lee (2022) according to which the effectiveness of AI systems is strongly dependent on the context of use, and that a 'one size fits all' approach is rarely effective in education.

Finally, on a practical level, the lack of effect of versatility calls for a certain caution in digital equipment policies. Rather than looking for universal solutions, establishments would benefit from selecting specialized tools by discipline, even if it means multiplying the interfaces, provided that each is perfectly adapted to the specific needs of the correctors. This approach, more modest in appearance, could prove more effective than a race for versatility, the real benefits of which remain to be demonstrated.

Strategic recommendations and research perspectives

The results of this research call for recommendations adapted to the institutional, infrastructural and human realities of the Algerian university context.

For institutions (ministries, rectorates, university management):

• **Start with a restricted and controlled scope:** rather than a generalized deployment, initiate pilot projects in sectors where standardized correction predominates (exact sciences, languages, MCQs), before gradually extending to disciplines with open answers**.**

• **Invest in lightweight and accessible infrastructures:** favor software solutions operating offline or with low bandwidth, taking into account connectivity constraints in certain regions (notably in Laghouat and Ouargla). Pool resources at the scale of regional centers**.**

• **Create local support units:** make up these units of teacher-trainers and technicians, trained by partners such as CERIST (Research Center on Scientific and Technical Information) or the ANR (National Research Agency), to ensure local monitoring**.**

• **Negotiate partnerships with educational software publishers and Algerian startups:** rather than importing solutions designed for other contexts, co-build tools adapted to local linguistic and educational specificities.

• **Include digital assessment in the national digital transformation strategy:** coordinate initiatives with the Ministry of Higher Education and Scientific Research to ensure overall consistency and sustainable funding.

For teaching teams:

- **Adopt progressive familiarization**: start by using AI for the correction of closed exercises or multiple choice questions, before exploring its potential for more complex productions.
- **Get involved in the design and evaluation of tools**: teachers are best placed to identify the real needs and blind spots of algorithms. Their feedback must fuel the development loops.
- **Develop hybrid skills**: integrate digital training modules into continuing training plans (in particular via the national training of trainers system), emphasizing the critical interpretation of the results generated by AI.
- **Establish inter-establishment communities of practice**: encourage exchanges of experiences between teachers from the universities of Laghouat, Ouargla and other establishments facing the same challenges.

For developers and solution designers:

•**Prioritize contextual adaptation**: rather than seeking technical sophistication, design simple interfaces, in Arabic and French, taking into account the work habits of Algerian teachers (low familiarity with complex digital environments).

•**Develop specialized disciplinary modules**: a single solution for all sectors is illusory. Separate modules for mathematics, literature, social sciences, etc., will better match educational expectations.

•**Involve users from the design phase**: integrate teacher-correctors into the steering committees for effective co-construction.

Guarantee the transparency and auditability of algorithms: to respond to ethical concerns and the requirements of Law 18-07 on the protection of personal data, teachers must be able to understand and challenge AI decisions.

For research (areas to be explored further):

Beyond this study, several projects deserve to be explored to shed light on the digital transition of evaluation in Algeria:

- **The differentiated impact according to the disciplines and types of exercises:** longitudinal studies comparing the sectors would make it possible to refine the recommendations.
- **The effects on the quality of learning and teaching practices**: does AI change the way of teaching and assessment? How does it affect the educational relationship?
- **The comparative effectiveness of training methods**: what format (face-to-face, hybrid, distance learning) and what content (technical, educational, ethical) allow the best appropriation of the tools by teachers?
- **Perception and acceptance by students**: the study did not integrate the learners' point of view, which is essential for a fair and formative evaluation.
- **Algorithmic biases in a multilingual and multicultural context**: is AI trained on predominantly Western corpora capable of fairly processing productions in Arabic or mixing several languages?

Conclusion

This research is based on a clear observation: artificial intelligence is expected to play a major role in optimizing correction chains within Algerian universities. But this potential, as promising as it may be, will only be realized under certain conditions. Because the effectiveness of these systems is not automatic: it is built, organized and negotiated with those on the ground.

Two pillars emerge from the results as the real keystones of successful integration. The first is analytical precision, which establishes the credibility of algorithmic feedback and makes it possible to reduce differences between correctors. The second is the ease of appropriation by teachers: without intuitive and rapid handling, the most sophisticated tool remains unused, especially in a context where administrative burdens leave little room for experimentation.

This dual requirement outlines operational priorities for establishments. Rather than giving in to the lure of plethora of platforms, they are recommended to select solutions already proven on these two criteria, and to invest massively in progressive support programs. Only a continuing training approach, adapted to the rhythms and needs of teachers, can transform a technical tool into a resource truly integrated into daily professional life.

In contrast to this observation, the study reveals that perceived versatility – often put forward as a commercial argument – has no significant impact on real efficiency. This result calls for double caution: on a conceptual level, it suggests reconsidering the way in which we define and measure this notion; on a practical level, it encourages calls for tenders to be refocused around a tight functional core, rather than on an escalation of superfluous modules. Developers are thus tacitly oriented towards a clear priority: ergonomics and contextual adequacy take precedence over technical abundance. This change of direction requires close collaboration with the teacher-correctors, who are the only ones able to adjust the interfaces to real uses and the organizational and cultural specificities of the Algerian university. Without this attention, the most successful innovations would encounter predictable resistance.

On the part of teachers, these technologies are seen as valuable aids to lighten the correction load. But our results say it strongly: human expertise remains irreplaceable. Automation cannot in any way replace educational judgment, otherwise it would deprive the evaluation of its substance. Therefore, training cannot be limited to technical familiarization. It must include in-depth reflection on the ways of articulating the contribution of the algorithm with professional gestures, in order to make it an ally and not a competitor.

Ultimately, this research does not just shed light on concrete and immediate choices. It also opens a larger project: that of the evolution of evaluative practices in the digital age. AI is gradually establishing itself as a lever for modernization, but on one demanding condition: that the university community seizes it not as a turnkey solution, but as an analyzer of its own practices. This involves continuing to closely observe uses, listening to feedback from the field, and continuously adjusting the systems to the needs actually expressed. Because it is at this price – that of attentive and iterative co-construction – that technological innovation will be able to keep its educational promises.

References:

Baker, R. S. (2023). Artificial intelligence in education: The critical questions. *Educational Technology Research*, 71(2), 45-62.

Benzane, A., & Khelfaoui, M. (2021). The challenges of massification in Maghreb universities : what digital solutions ? International Journal of Educational Sciences, 15(2), 45-62.

Bouguelaa, S. (2022). Artificial intelligence for the evaluation of learning: potential and limits in the Algerian educational context. University Editions.Bourdieu, P., & Passeron, J. C. (1990). *Reproduction in education, society and culture*. Sage Publications.

Chen, L., Wang, M., & Liu, Y. (2023). Adaptive learning systems: A systematic review of the literature. *Computers & Education*, 189, 104-123.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1), 1-35.

European Commission. (2023). *Ethical guidelines on the use of artificial intelligence in education*. Publications Office of the European Union.

Holmes, W., Porayska-Pomsta, K., & Holstein, K. (2023). Ethics of AI in education: State of play and potential futures. *International Journal of Artificial Intelligence in Education*, 33(4), 89-120.

Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 6300-6308.

Kew, S. N., & Tasir, Z. (2024). Learning analytics in higher education: A systematic literature review. *Journal of Computing in Higher Education*, 36(1), 1-25.

Khelfaoui, H. (2020). Barriers to the adoption of ICT by university teachers in Algeria: between cultural resistance and training deficit. In ICT and Educational Transformation (pp. 87-104). Presses Universitaires du Maghreb.

Kim, S., & Lee, J. (2024). Educational data mining: Applications and trends. *Computers & Education*, 192, 104-123.

Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. *Ethics in Artificial Intelligence*, 1-38.

Li, X., & Wang, H. (2024). Automated essay scoring: A review of the literature. *Journal of Educational Computing Research*, 62(3), 45-67.

Park, Y., & Chen, J. (2024). Predictive analytics in education: A review of the literature. *Educational Technology Research and Development*, 72(1), 89-112.

Rodriguez, M. U., & Garcia, P. (2024). Natural language processing in education: Current trends and future directions. *Computers & Education*, 195, 104-123.

Selwyn, N. (2020). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.

Siemens, G., & Baker, R. S. (2023). Learning analytics and educational data mining: An updated review. *Journal of Learning Analytics*, 10(1), 1-15.

Tala-Ighil, N. (2019). Stakeholders' perceptions and resistance to digital evaluation. Case of Algerian universities. Cahiers de la Recherche en Pédagogie, 12(1), 88-105.

Thompson, K., & Davis, A. (2024). Early warning systems in higher education: A systematic review. *Journal of Educational Technology Systems*, 52(2), 145-167.

UNESCO. (2023). *AI and education: Guidance for policy-makers*. United Nations Educational, Scientific and Cultural Organization.

Vaswani, A., Shazeer, N., & Parmar, N. (2023). Transformer models in education: A technical review. *IEEE Transactions on Learning Technologies*, 16(3), 45-62.

Wang, F., & Liu, Y. (2024). Personalized learning in the AI era: A comprehensive review. *Educational Technology Society*, 27(1), 89-104.

Zawacki-Richter, O., & Marin, V. I. (2023). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 20(1), 1-24.

Zhang, H., & Aslan, A. B. (2024). AI technologies for education: Recent advances and future directions. *Computers & Education*, 190, 104-123