Generative Artificial Intelligence as Evaluative Scaffolding: Effects on Students' Perceptions and Dispositions in an Authentic Learning Task

Beatrice Doria
Department of Humanities, Education and Sport, Pegaso University, Italy
beatrice.doria@pegaso.it
ORCID: 0000-0002-3894-9460

Giorgia Slaviero
Department of FISPPA, University of Padova, Italy
giorgia.slaviero.1@phd.unipd.it
ORCID: 0009-0000-9312-4683

Cristina Zaggia
Department of FISPPA, University of Padova, Italy
cristina.zaggia@unipd.it

Valentina Grion
Department of Humanities, Education and Sport, Pegaso University, Italy
valentina.grion@pegaso.it
ORCID: 0000-0002-2051-1313

ABSTRACT
In recent years, the integration of generative artificial intelligence (GenAI) into educational contexts has raised important pedagogical questions, particularly regarding its role in authentic assessment processes and the development of evaluative judgement. This study explores the integration of GenAI within an experiential learning and authentic assessment device centered on the construction, revision, and self-assessment of rubrics, investigating how this experience is associated with changes in the perceptions and dispositions of university students who are future primary school teachers. Using a single-group pre-post quasi-experimental design, 144 participants constructed an assessment rubric and subsequently revised it using both an exemplar and GenAI tools, in a sequence consistent with Kolb's experiential learning cycle. Three dimensions were measured before and after the intervention: reflective and criterion-based approaches, attitudes towards AI, and technostress/computer anxiety. Results, analyzed through Wilcoxon signed-rank tests and non-parametric correlations, show statistically significant changes across all three dimensions: an increase in reflective and criterion-based approaches ($r = -0.725$), a marked increase in positive attitudes towards AI ($r = -0.890$), and a reduction in technostress ($r = 0.421$). Lower initial levels across the dimensions were associated with larger changes, while no significant differences emerged related to socio-demographic characteristics or the tool perceived as most useful. Findings suggest that GenAI can function as genuine evaluative scaffolding — an agent-to-learn-with rather than a substitute for human judgement — capable of fostering critical reflection and the formation of evaluative judgement, provided it is integrated into intentionally designed and pedagogically mediated learning environments.

INTRODUCTION
In recent years, higher education has undergone profound transformations that are significantly redefining how learning occurs. In this rapidly evolving scenario, the emergence of generative artificial intelligence (GenAI) represents one of the most disruptive phenomena, as it introduces new educational opportunities while at the same time raising important pedagogical, epistemological, and ethical issues. The ability of these systems to produce complex and seemingly "human" content has fuelled a broad academic debate, particularly regarding issues such as integrity, the validity of assessment practices, the redefinition of authorship, and the development of higher-order cognitive skills (Chiu, 2023; Crawford et al., 2023; Stokel-Walker & Van Noorden, 2023; Baidoo-Anu & Owusu Ansah, 2023). At the same time, the growing diffusion of GenAI is calling into question the epistemological foundations of traditional university assessment, as tasks historically used to check knowledge or procedural execution can today be at least partially delegated to AI systems. This transformation makes it increasingly clear that there is a need to design assessment practices capable of valuing complex cognitive processes, reflection, judgement, and authentic performance, rather than limiting themselves to the simple evaluation of the final product.

Considering these transformations, traditional models of teaching and assessment – historically based on transmissive and reproductive logics of knowledge – are becoming progressively inadequate in relation to contemporary educational goals. The literature shows that conventional assessment practices, often based on standardized tests and predominantly summative logics, are no longer fully aligned with the need to develop critical, reflective, and problem-solving competences in complex and situated contexts (Wiggins, 1990; Villarroel et al., 2018; Bearman et al., 2022). In response to this misalignment between assessment practices and educational aims, the paradigm of *authentic assessment* has progressively emerged as a perspective able to valorise student performance in meaningful tasks anchored in real or plausible contexts and oriented towards the integrated mobilisation of knowledge, skills, and judgement (Koh, 2017; Ajjawi et al., 2023; Ashford-Rowe et al., 2014). From this perspective, assessment ceases to be understood as a simple measurement of learning and instead becomes a dynamic and participatory process, closely intertwined with the learning experience and oriented towards the construction of meaning (Boud & Soler, 2016).

Within this reconceptualization of assessment, the student's role becomes increasingly central. Practices such as the collaborative construction of assessment rubrics show how assessment can itself become a learning activity, insofar as students are invited to negotiate criteria, make explicit quality standards, and develop evaluative judgement (Panadero & Jonsson, 2013; Tai et al., 2018). In this sense, the co-construction of rubrics is not only a technical assessment tool but also a metacognitive and epistemic device through which students learn to recognize quality, compare performances, and internalize criteria that are useful for self-regulation and lifelong learning (Boud et al., 2018; Tai et al., 2018). This outlines a vision of assessment as a formative and participatory process, closely connected to the development of self-regulation, feedback literacy, and metacognitive awareness (Zimmerman, 2002; Panadero, 2017).

This perspective finds solid theoretical grounding in Kolb's (1984) experiential learning model. According to this model, learning develops through a recursive cycle that integrates concrete experience, reflective observation, abstract conceptualization, and active experimentation (Kolb, 1984; Kolb & Kolb, 2018), highlighting the central role of experience and reflection in knowledge construction (Moon, 2004). It is therefore no coincidence that the literature reveals a strong convergence between experiential learning and *authentic assessment*: both approaches place at the center the student's activity, the relevance of proposed tasks, and the development of transferable and contextualized competences (Geerling et al., 2023; Fry et al., 2023). Within this integrated theoretical framework, the experiment described here is structured according to a logic consistent with Kolb's cycle: individual construction of the rubric represents the phase of concrete experience; comparison with exemplars and GenAI activates reflective observation; critical revision of the rubric corresponds to abstract conceptualization; finally, self-assessment closes the cycle through active experimentation.

It is precisely within this convergence between authentic assessment and experiential learning that the transformative potential of GenAI is located. Whereas traditional educational technologies have predominantly served to facilitate access to information, generative artificial intelligence can now intervene actively in learning processes, contributing to the construction of experiences, reflection, and evaluation (Zawacki-Richter et al., 2019; Luckin et al., 2016). Recent studies suggest that these tools can support experiential learning through the generation of realistic scenarios, the facilitation of reflection, and the provision of personalized and continuous feedback (Ifenthaler & Schumacher, 2023; Escalante et al., 2023; Kasneci et al., 2023). In particular, Salinas-Navarro et al. (2024) propose a systematic integration between GenAI, experiential learning, and authentic assessment, highlighting how these technologies can support all phases of Kolb's cycle and foster the development of complex competences. In line with this perspective, in the present experiment GenAI is combined with exemplars as a form of evaluative scaffolding: both function as external resources for improving the rubric, although they differ in interaction modes, the level of personalization of feedback, and the type of cognitive processing required of students.

From this standpoint, GenAI is not configured simply as a support tool but as a genuine agent-to-learn-with (Salinas-Navarro et al., 2024), capable of stimulating critical thinking, reflection, and situated action (Jonassen, 1996; Kim & Baylor, 2006). However, precisely this growing integration between technology, learning, and assessment also highlights a relevant theoretical and empirical gap. Most studies focus on the use of GenAI as a support for content production or automated feedback provision, while contributions that systematically address the issue of instructional and assessment design mediated by artificial intelligence – particularly in authentic assessment and experiential learning contexts – remain limited (Salinas-Navarro et al., 2024; Zawacki-Richter et al., 2019). Similarly, research on the use of rubrics and AI in assessment processes is still emerging and fragmented, especially with respect to the construction of evaluative judgement and human–AI interaction dynamics in authentic educational contexts (Demir & Çüm, 2026). This gap is particularly significant if we consider that the effectiveness of GenAI integration does not depend solely on its technological characteristics but results from a complex interplay between pedagogical, design-related, and individual dimensions. Recent studies show that generative AI systems can produce superficial feedback, hallucinatory information, or foster forms of cognitive dependence and passivity, especially in the absence of adequate pedagogical mediation and human supervision (Jia et al., 2024; Demir & Çüm, 2026). It follows that the educational value of GenAI does not lie simply in the automation of processes but rather in the quality of instructional design and the capacity to critically integrate artificial support within reflective and formative practices.

From this perspective, alongside teachers' competences – whose centrality is widely documented in the literature on faculty development, assessment literacy, and the TPACK framework (Mishra & Koehler, 2006; Koehler & Mishra, 2009) – the individual characteristics of students interacting with such tools also become crucial. Recent literature highlights how cognitive, emotional, attitudinal, and socio-demographic factors can significantly influence how GenAI is accepted, used, and integrated in educational contexts (Symasek et al., 2025). However, the role of individual differences in AI-mediated authentic assessment processes remains scarcely explored.

The present study aims to explore the integration of GenAI within an authentic assessment and experiential learning device centered on the construction, revision, and self-assessment of rubrics. Specifically, the research investigates the role of GenAI as a scaffold for the development of evaluative judgement and analyses how individual student differences influence perceptions, interaction modalities, and use of AI in reflective assessment processes. From this perspective, this study does not view GenAI merely as a technical or productive tool, but as a potential facilitator of reflective and assessment processes, capable of supporting the formation of judgement, critical engagement with criteria, and the development of evaluative judgement within authentic assessment practices. By integrating authentic assessment, experiential learning, and human–AI interaction, the study intends to contribute to a still under-explored research area related to the pedagogical design of assessment practices supported by artificial intelligence in higher education.

INDIVIDUAL DIFFERENCES AND ACCEPTANCE OF ARTIFICIAL INTELLIGENCE IN EDUCATIONAL CONTEXTS

The integration of Artificial Intelligence in educational contexts cannot be interpreted solely in terms of technological innovation or redesign of instructional and assessment devices. Rather, it must be understood as a complex, multi-level process that is deeply influenced by the characteristics of the actors involved. As anticipated in the introduction, alongside the pedagogical perspectives of authentic assessment and experiential learning, there is an urgent need to consider the role of individual differences in processes of acceptance, use, and adoption of AI-based technologies.

Recent literature underscores that individual differences – understood as the set of relatively stable psychological, cognitive, emotional, and socio-demographic characteristics – constitute a decisive variable in shaping the relationship between individuals and technologies (Symasek et al., 2025; Kelly et al., 2023). This perspective helps explain why, despite equivalent opportunities offered by GenAI, we observe highly heterogeneous levels of use, appropriation, and integration in teaching and assessment practices. In this direction, classical theoretical models of technology acceptance – such as the Technology Acceptance Model (TAM; Davis, 1989), the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003), and the Theory of Planned Behaviour (TPB; Ajzen, 1991) – offer a consolidated interpretive basis but remain partial, as they tend to privilege general cognitive variables such as perceived usefulness and ease of use without fully integrating the subjective and differential dimension of users (Symasek et al., 2025; Crompton & Burke, 2023).

A first connecting element among these perspectives concerns the distinction between acceptance and adoption. While acceptance refers mainly to attitudes and intentions towards technology, adoption involves its concrete and continuous use in daily practices (Davis, 1989; Venkatesh et al., 2003). In university settings, this distinction is particularly relevant, since declared willingness to use GenAI does not automatically guarantee its effective integration in teaching, learning, and assessment processes (Crompton & Burke, 2023). It is precisely in the gap between intention and action that individual differences play a decisive role, influencing both the formation of attitudes towards AI and their translation into concrete educational practices (Symasek et al., 2025; Ajzen, 1991).

Among the most relevant factors, self-efficacy emerges, that is, the perception of one's own ability to use technology effectively. Numerous studies show that self-efficacy is positively associated with both perceived ease of use and intention to use digital technologies (Bandura, 1997; Compeau & Higgins, 1995; Zhang et al., 2017), suggesting that individuals with higher confidence in their digital competences are more likely to experiment with AI tools and integrate them more consistently into learning activities. This dimension appears closely linked to prior experience, which helps reduce perceived barriers and facilitate technological appropriation processes (Agarwal & Prasad, 1999; Varma & Marler, 2013). In the context of the present experiment, this variable is particularly relevant, as students are asked to work with GenAI in a structured and relatively novel task – the construction and revision of an assessment rubric – that requires technological competence, reflective capacity, and pedagogical judgement at the same time.

In parallel, trust represents another key element in AI acceptance processes. The literature shows that trust in technological systems – understood in terms of perceived reliability, transparency, and security – helps reduce perceived risk and foster intention to use (McKnight et al., 2002; Wu & Chen, 2005; Faqih, 2011). This aspect is even more relevant in the case of GenAI, characterized by high complexity and algorithmic opacity, which can generate uncertainty, resistance, or ambivalent attitudes in users (Symasek et al., 2025). In the absence of adequate levels of trust, even technologies perceived as useful risk being used superficially, inconsistently, or systematically avoided.

These dimensions are accompanied by emotional and attitudinal factors. Positive attitudes towards technology are generally associated with a greater propensity to use it (Ajzen, 1991; Venkatesh et al., 2003), whereas conditions such as technostress – defined as the negative response arising from the inability to adapt in a balanced way to new technologies (Ragu-Nathan et al., 2008) – or computer anxiety can significantly hinder technology adoption processes (Thatcher & Perrewé, 2002; Rosen & Weil, 1995; Joo et al., 2016). As highlighted by Symasek et al. (2025), these variables show a systematic negative relationship with intention to use and perceived usefulness of digital tools, whereas technological self-efficacy appears to act as a positive moderator. This shows that the introduction of GenAI in educational contexts requires not only technical competences but also explicit attention to users' emotional and psychological dimensions.

Another layer of complexity is represented by cognitive and learning styles. Research suggests that a preference for more systematic or intuitive modes of information processing – typically framed as analytical vs. holistic styles – influences both attitudes towards technological innovation and the ways of interacting with AI systems (Chae et al., 2020; Chakraborty et al., 2008; Riding & Rayner, 1998). Likewise, personality traits associated with the Big Five model show significant associations with technology adoption processes: openness to experience tends to favor willingness towards innovation, whereas high levels of neuroticism are associated with greater anxiety and distrust towards AI (Svendsen et al., 2013; Sindermann et al., 2022). These findings reinforce the idea that GenAI adoption is not a uniform process but a deeply situated, relational, and differentiated phenomenon.

Socio-demographic variables, particularly gender and age, also yield complex and sometimes contradictory results. Regarding gender, some studies report higher levels of computer anxiety among women (Rosen & Weil, 1995; Dos Santos & Santana, cited in Symasek et al., 2025), although such differences tend to progressively decrease with increased digital experience and exposure to technologies. On the generational level, research questions the assumption that so-called "digital natives" automatically possess superior technological competences. Symasek et al. (2025), for example, find no significant age-related differences in the adoption of specific AI technologies, suggesting that factors such as motivation, prior experience, and context of use carry more weight than simple generational membership.

Overall, the literature highlights substantial continuity between the factors influencing technology acceptance in general and those that regulate AI adoption. However, the distinctive nature of generative artificial intelligence – characterized by autonomous capabilities, content production, conversational simulation, and significant ethical implications – requires an extension of traditional interpretive models (Kasneci et al., 2023; Floridi et al., 2018). As Symasek et al. (2025) emphasize, technology acceptance research offers a necessary but not sufficient theoretical basis for fully understanding AI adoption dynamics. Dimensions such as algorithmic trust, perceived agency, system transparency, and ethical implications of generative tools introduce constructs that classic models do not yet fully capture (Dignum, 2019).

In relation to this study's objectives, these considerations are strategically relevant. If GenAI is understood as an agent-to-learn-with (Salinas-Navarro et al., 2024) within experiential learning and authentic assessment environments, then its effectiveness will depend not only on the quality of instructional design but also on the capacity to respond to diverse user profiles (Symasek et al., 2025; Panadero, 2017). This implies the need to integrate technological, pedagogical, and individual dimensions within a unified framework, in which deliberate training plays a central role in supporting teachers and students in developing critical, reflective, and evaluative competences (Koehler & Mishra, 2009; Laurillard, 2012; Steinert et al., 2016).

From this perspective, taking individual differences into account is not an accessory element but a necessary condition to promote GenAI integration that is truly effective, sustainable, and consistent with experiential learning and authentic assessment principles (Kolb, 1984; Ajjawi et al., 2023). This need is particularly relevant because the AI acceptance of literature is still limited and fragmented, especially regarding the role of individual differences in authentic educational and assessment contexts. As recent reviews show, research still needs to examine in greater depth how factors such as self-efficacy, trust, experience, attitudes, and technological anxiety concretely influence interaction with AI in learning and assessment processes (Symasek et al., 2025). The present study contributes to filling this under-explored area by investigating how students' individual differences influence interaction with GenAI tools in authentic and reflective assessment practices.

AIM

This study aims to explore the perceptions and dispositions of university students, future primary school teachers, regarding the use of generative artificial intelligence (GenAI) as a reflective and evaluative support tool within an authentic activity of constructing, revising, and self-assessing rubrics. In line with Symasek et al. (2025), who highlight that "Future research should consider incorporating these ID factors into studies on AI technology acceptance and adoption to deepen our understanding of their role" (p. 26), the present research intends to contribute to a deeper understanding of the role of individual difference factors (ID factors) in AI-based technology acceptance and adoption processes, exploring how differing individual dispositions influence students' perceptions, attitudes, and interaction modalities with GenAI tools in authentic assessment and experiential learning contexts.

In particular, the study aims to analyze whether, and to what extent, participation in a GenAI-mediated instructional experience is associated with changes in three dimensions measured before and after the activity:

1. Reflective and criterion-based approaches.
2. Attitudes and dispositions towards artificial intelligence.
3. Technostress and computer anxiety.

The study also aims to examine the magnitude and individual variability of the observed changes, considering not only group-level mean variations but also the distribution of individual changes among participants. From this perspective, the analysis intends to understand how students with different initial dispositions respond to GenAI integration in authentic assessment and experiential learning practices.

Finally, the study seeks to investigate whether possible differences in observed changes and post-intervention perceptions are associated with: (1) students' individual characteristics; (2) the tool perceived as most useful during the rubric revision activity.

Based on these objectives, the following research questions were formulated:

- **RQ1.** Are there changes in students' dispositions between pre-test and post-test in the dimensions of learning style, attitudes towards AI, and technostress?
- **RQ2.** To what extent did students' scores change from pre-test to post-test across the different dimensions, and were these changes statistically and practically significant?
- **RQ3.** Are the initial levels in the considered dimensions associated with the magnitude of observed change after the experience with GenAI?
- **RQ4.** Are there differences in the observed changes and in post-intervention perceptions in relation to:
  a) students' individual characteristics;
  b) the tool perceived as most useful during the activity?

DESIGN, CONTEXT, AND INSTRUMENTS

This study adopts a single-group pre–post quasi-experimental design to explore students' perceptions and dispositions regarding the integration of generative artificial intelligence (GenAI) into an authentic assessment activity focused on the construction and revision of rubrics. Consistent with an exploratory perspective, the research aims to analyze changes in students' dispositions before and after participation in the laboratory experience, without assuming direct causal relationships between GenAI use and the observed changes.

The instructional intervention was designed according to principles of experiential learning and authentic assessment, involving students in tasks of designing, revising, and self-assessing assessment tools. In this context, GenAI was used as a reflective and dialogic support resource for improving the produced rubric, within a collaborative activity mediated by exemplars and peer discussion.

The study took place within a university laboratory for students enrolled in a Primary Teacher Education degree program. Participation occurred during regular curricular teaching activities and was not associated with summative final assessment.

EXPERIMENTAL PROCEDURE

The experiment was structured in four sequential phases, designed to actively involve students in the construction, revision, and self-assessment of rubrics using exemplars and GenAI tools in combination.

Phase 0 – Initial Survey

Before starting the laboratory activity, participants completed a pre-intervention questionnaire lasting approximately 7–10 minutes. The questionnaire aimed to collect socio-demographic data and initial dispositions regarding interaction with digital technologies and artificial intelligence tools.
Specifically, three dimensions were investigated:

a) reflective and criterion-based approaches in learning and decision-making processes;
b) attitudes towards artificial intelligence;
c) technostress and computer anxiety.

Phase 1 – Construction and Submission of the Rubric

Subsequently, students were involved in the collaborative construction of an assessment rubric, working in small groups. The activity included:
a) identifying a competence to be assessed, selected based on the National Curriculum Guidelines for school;
b) defining the relevant instructional context;
c) identifying assessment criteria.

Based on these elements, participants developed a rubric structured in four performance levels. To support the design process, an exemplar ("rubric of the rubric") was provided, constructed on Brookhart's (2013) framework, and containing criteria and indicators useful for building effective and coherent assessment rubrics.

Phase 2 – Revision and Self-Assessment

In the next phase, groups revised the produced rubric using the provided exemplar and generative artificial intelligence tools.

GenAI was used as a reflective support to obtain feedback, revision suggestions, and clarifications on the quality of the constructed rubric. To guide interaction with the system, participants were asked to share the previously provided exemplar with the AI, requesting feedback aligned with the stated assessment criteria.

Groups could engage in iterative interaction with the AI, formulating further requests for clarification, elaboration, or revision of the received feedback. The activity was not intended to delegate the evaluative task to the AI system, but rather to promote processes of critical comparison, revision, and meta-evaluative reflection.

At the end of this phase, each group submitted the final revised version of the rubric.

Phase 3 – Final Survey

At the end of the laboratory experience, participants completed a post-intervention questionnaire aimed at detecting:

a) any changes in initial dispositions;
b) perceptions regarding the use of GenAI;
c) perceived usefulness of the tools employed;
d) perceived development of evaluative and reflective competences.

DATA COLLECTION INSTRUMENTS

Data were collected through two structured questionnaires, mirrored in content, administered to each participant before (pre-intervention) and after (post-intervention) the experiment (Appendix A).
The initial questionnaire was developed following the theoretical framework proposed by Symasek et al. (2025), which highlights the role of individual differences in the acceptance and adoption of artificial intelligence technologies. In line with this model, the instrument collected socio-demographic variables (gender, age, generational group) and psychological and dispositional dimensions relevant to interaction with advanced digital technologies.

The dimensions analyzed were measured using multi-item scales for a total of 19 items, structured as follows:

- Reflective and criterion-based approaches in learning and decision-making processes (5 items: Q5–Q9)
  This scale assesses individual preferences in learning processes and decision-making modes in evaluative contexts, with particular attention to the use of criteria, models, and reflective elements.
- Attitude towards artificial intelligence (5 items: Q10–Q14)

This scale measures the degree of openness, interest, and perceived usefulness of AI in educational and assessment contexts.

- Technostress and computer anxiety (5 items: Q15–Q19)
  This dimension captures emotional and cognitive responses associated with the use of digital technologies, including aspects of anxiety, perceived stress, and adaptability.

All items were measured using a six-point self-anchoring Likert scale (1 = "not at all true for me"; 6 = "completely true for me"), chosen to reduce the tendency towards neutral responses and ensure good sensitivity of the instrument.

The post-intervention questionnaire included an additional section to capture students' perceptions of the experience and of GenAI use in the rubric construction and revision process. Specifically, information was collected on: (a) perceived usefulness of artificial intelligence; (b) usefulness of exemplars; (c) perceived development of evaluative competences; (d) perceived future role of AI in instructional and assessment design; (e) perceived contribution of the tools used to understanding the assessment process. Participants were also asked to indicate which tool they perceived as most useful for improving the produced rubric.

DATA ANALYSIS PROCEDURES

Statistical analyses were conducted using jamovi software (version 2.6), following a multi-step procedure aimed at describing the sample, exploring variable distributions, and verifying observed changes between pre- and post-intervention surveys.

First, descriptive analyses of socio-demographic variables (gender, age, generational group, and professional experience) were carried out by calculating absolute frequencies, percentages, means, and standard deviations, to outline the sample profile.

Next, for the questionnaire-based dimensions – learning and decision-making style, attitude towards artificial intelligence, and technostress/computer anxiety – composite scores were computed as the mean of items belonging to each scale. For each dimension, in both time points (pre and post), descriptive statistics (mean, median, standard deviation, minimum and maximum values) were calculated to explore score trends and data distribution.

Normality of distributions was assessed using the Shapiro–Wilk test, applied both to composite scale scores and to difference variables calculated as Δ = post score − pre score. Since several distributions deviated significantly from normality, inferential analyses were conducted primarily using non-parametric procedures.

To examine changes between pre-test and post-test, difference variables were constructed for each dimension (Δ learning style, Δ attitude towards AI, Δ technostress). Differences between the two measurement points were analyzed using the Wilcoxon signed-rank test for paired samples. For each comparison, effect size was estimated through rank-biserial correlation (r), in order to gauge the practical magnitude of observed differences.

To further investigate individual variability in changes, correlational analyses were subsequently conducted between socio-demographic variables, initial scale scores, and difference variables. Given the ordinal nature of some variables and the non-normality of distributions, Spearman's ρ and Kendall's Tau-b (τb) correlations were primarily used. Pearson correlations were reported only for associations between continuous variables that met parametric assumptions. These analyses allowed exploration of associations between individual characteristics, initial dispositions, and the magnitude of changes after the GenAI experience.

To better understand students' perceptions of the experience, descriptive analyses were also performed on variables collected exclusively in the post-intervention questionnaire, concerning perceived usefulness of AI, development of evaluative competences, AI's future role in assessment design, and the contribution of the tools used to understanding the assessment process.

Finally, to explore potential differences in observed changes relative to the tool perceived as most useful for rubric improvement (AI, exemplar, both, or neither), comparisons between independent groups were conducted using the non-parametric Kruskal–Wallis test. This analysis allowed verification of possible differences in change scores and post-intervention perceptions among different profiles of tool use.

For all analyses, the statistical significance level was set at $p < .05$.

RESULTS

SCALE RELIABILITY

Internal consistency of the scales used was assessed through Cronbach's alpha and McDonald's omega, considered complementary indices of reliability for psychometric measures.

The scale for *reflective and criterion-based approaches in decision-making processes* showed modest reliability levels ($\alpha = .50$; $\omega = .52$), suggesting moderate internal consistency among items. Corrected item–total correlations ranged between .20 and .37. These values may reflect the multidimensional and exploratory nature of the construct, which integrates aspects related to reflective processes, use of criteria, and decision-making modes in assessment contexts.
The scale for *attitude towards artificial intelligence* displayed high internal reliability ($\alpha = .90$; $\omega = .90$), indicating excellent consistency across items. Corrected item–total correlations ranged from .71 to .80, indicating a homogeneous contribution of indicators to the construct.

The *technostress/computer anxiety* scale also showed satisfactory reliability ($\alpha = .77$; $\omega = .79$). In preliminary analysis, item Q15 was reverse-scored, as it was phrased in the opposite direction to the other indicators. After reversal, all item–total correlations were positive (range: .36–.70), indicating adequate alignment of items with the underlying construct.

Overall, the analyses support good reliability for the attitude towards AI and technostress scales, whereas the reflective and criterion-based approaches scale exhibits lower internal consistency and therefore requires some caution in interpretation.

SAMPLE DESCRIPTION

The sample consists of 144 participants. The gender distribution shows a clear prevalence of women (n = 139; 96.5%) compared to men (n = 5; 3.5%).

Considering the generational group, most participants belong to Generation Z (n = 120; 83.3%), followed by Millennials (n = 22; 15.3%), while a small number indicated "Other/Do not know" (n = 2; 1.4%).

Regarding the professional experience, 110 participants (76.4%) reported not currently working in school settings. Among those employed, 19 (13.2%) work in primary school, 12 (8.3%) in preschool, and 3 (2.1%) in other educational contexts.

In the post-intervention questionnaire, students were also asked which tool they perceived as most useful for improving the rubric. Over half of participants indicated GenAI as the main support (n = 79; 54.9%), while 61 students (42.4%) reported finding the combined use of AI and exemplars most useful. Only a small proportion indicated only the exemplar (n = 3; 2.1%) or neither tool (n = 1; 0.7%).

DESCRIPTIVE CHANGES IN STUDENTS' DISPOSITIONS BETWEEN PRE-TEST AND POST-TEST (RQ1)

Descriptive analyses were conducted on mean scores for the three questionnaire dimensions – learning and decision-making style, attitude towards AI, and technostress/computer anxiety – measured at pre- and post-intervention. Analyses were carried out on a sample of 144 participants, with no missing data.

Overall, results show variations in mean scores between the two measurements (Table 1).

Table 1. *Descriptive statistics of variables in pre and post measurements*

Variable	N	M Pre	SD Pre	Md Pre	M Post	SD Post	Md Post
Reflective and criterion-based	144	4.78	0.54	4.80	5.06	0.52	5.00
Attitude towards AI	144	4.12	1.05	4.20	4.84	0.80	5.00
Technostress	144	3.12	1.01	3.20	2.90	0.95	3.00

Note. Six-point Likert scale (1 = "not at all true for me"; 6 = "completely true for me").

Scores for reflective and criterion-based approaches show an increase from pre-test (M = 4.78; SD = 0.54) to post-test (M = 5.06; SD = 0.52). Attitude towards AI similarly increases, from M = 4.12 (SD = 1.05) at pre-test to M = 4.84 (SD = 0.80) at post-test.

By contrast, technostress/computer anxiety shows a decrease in mean scores from pre-test (M = 3.12; SD = 1.01) to post-test (M = 2.90; SD = 0.95), indicating lower average levels of perceived discomfort associated with digital technologies after the laboratory experience.

Overall, standard deviations are slightly lower in the post-intervention measurement, suggesting reduced score dispersion and greater convergence of students' perceptions at the end of the activity.

Shapiro–Wilk normality tests indicated significant deviations from normality for some variables, particularly for attitude towards AI (Table 2).

Table 2. *Shapiro–Wilk normality test for pre and post variables*

Variable	W	p
STILE_PRE	0.977	.017
STILE_POST	0.970	.003
ATT_AI_PRE	0.955	< .001
ATT_AI_POST	0.930	< .001
TECHNO_PRE	0.984	.104
TECHNO_POST	0.986	.137

Note. p < .05 indicates significant deviation from normality.

Given these deviations, subsequent inferential analyses primarily employed non-parametric tests.

CHANGES BETWEEN PRE-TEST AND POST-TEST: MAGNITUDE, STATISTICAL SIGNIFICANCE, AND EFFECT SIZE (RQ2)

To examine the extent to which students' scores changed between pre- and post-test, difference variables (Δ = post − pre) were calculated for each dimension (reflective and criterion-based approaches, attitude towards AI, and technostress/computer anxiety). Descriptive statistics indicate positive mean changes for reflective and criterion-based approaches (M = 0.29; SD = 0.41; Md = 0.20) and, most notably, for attitude towards AI (M = 0.72; SD = 0.75; Md = 0.60). In contrast, technostress/computer anxiety showed a negative mean difference (M = −0.22; SD = 0.57; Md = −0.20), indicating a reduction in perceived discomfort related to digital technologies (Table 3).

Table 3. *Descriptive statistics for difference variables (Δ = post − pre)*

Variable	M	Md	SD	Min	Max
Δ Learning style	0.29	0.20	0.41	−0.80	1.40
Δ Attitude towards AI	0.72	0.60	0.75	−1.00	3.00
Δ Technostress	−0.22	−0.20	0.57	−1.80	1.20

Note. Positive values indicate an increase in post-intervention; negative values indicate a decrease relative to pre-test.

Although all three dimensions exhibited changes in the expected direction, the observed variability suggests that the magnitude of change differed across participants, indicating heterogeneous individual responses to the GenAI experience.

To determine whether these changes were statistically significant, Wilcoxon signed-rank tests for paired samples were conducted, as Shapiro–Wilk tests indicated deviations from normality. Significant pre–post differences emerged for all three dimensions (Table 4). Specifically, students reported greater use of reflective and criterion-based approaches to learning and decision-making (W = 1033, p < .001), a more positive attitude towards AI (W = 439, p < .001), and lower levels of technostress/computer anxiety following the intervention (W = 5774, p < .001).

Table 4. *Pre–post comparisons via Wilcoxon test*

Variable	W	p	r
Learning style	1033	< .001	−0.725
Attitude towards AI	439	< .001	−0.890

Technostress	5774	< .001	0.421

Note. r = rank-biserial correlation (effect size).

Effect sizes indicate that the intervention produced substantial practical changes in two of the three dimensions. Large effects were observed for learning style ($r = -0.725$) and, especially, attitude towards AI ($r = -0.890$), whereas the reduction in technostress/computer anxiety was associated with a moderate effect ($r = 0.421$). Overall, these findings suggest that the GenAI experience was associated with statistically significant improvements across all dimensions, with particularly pronounced effects on students' reflective learning approaches and attitudes towards AI.

ASSOCIATION BETWEEN INITIAL LEVELS AND CHANGES AFTER THE GENAI EXPERIENCE (RQ3)

To explore whether initial levels on the considered dimensions were associated with the magnitude of change following the GenAI experience, correlational analyses were conducted between pre-test scores and the corresponding difference variables (Δ = post − pre). Given non-normal distributions, Spearman correlations were used.

Results show statistically significant negative associations between initial scores and their corresponding difference variables (Table 5).

Table 5. *Correlations between initial scores and difference variables*

Variables	Spearman's ρ	p
APPROACH_PRE – Δ APPROACH	−0.421	< .001
ATT_AI_PRE – Δ ATT_AI	−0.663	< .001
TECHNO_PRE – Δ TECHNO	−0.355	< .001
Δ ATT_AI – Δ TECHNO	−0.241	.004

In particular, there is a strong negative correlation between initial attitude towards AI and the corresponding difference variable ($\rho = -0.663$; $p < .001$). This indicates that less favorable initial attitudes towards AI are associated with larger positive changes at post-test.

A similar pattern emerges for reflective and criterion-based approaches ($\rho = -0.421$; $p < .001$), suggesting that lower initial levels are associated with greater changes between pre- and post-test.

Regarding technostress/computer anxiety, the significant negative correlation ($\rho = -0.355$; $p < .001$) shows that higher initial levels of technological discomfort are associated with larger reductions after the experience.

Finally, there is a significant negative correlation between change in attitude towards AI and change in technostress ($\rho = -0.241$; $p = .004$), indicating that increases in favorable dispositions towards AI are associated with reductions in perceived technological discomfort.

These associations, however, must be interpreted with caution, as correlations between initial scores and difference scores can be partially influenced by statistical phenomena such as regression to the mean and by mathematical dependence between variables.

DIFFERENCES IN OBSERVED CHANGES IN RELATION TO INDIVIDUAL CHARACTERISTICS AND TOOLS USED (RQ4)

Possible differences in observed changes were then examined in relation to some individual characteristics and to the tool perceived as most useful during rubric revision. Given non-normal distributions and the categorical nature of variables, group comparisons were performed using the non-parametric Kruskal–Wallis test (Table 6).

Table 6. *Group comparisons for change scores and post-intervention perceptions*

Variable	χ^2	df	p	ε^2
DIFF_STILE	3.51	3	.319	.024
DIFF_ATT_AI	3.23	3	.358	.023
DIFF_TECHNO	3.77	3	.287	.026
Evaluative competences	2.84	3	.416	.020

Tools and improvement	2.93	3	.402	.021

Note. Kruskal–Wallis comparisons relative to the tool perceived as most useful (AI, exemplar, both, none).

With respect to generational group, no statistically significant differences were found in change scores for learning and decision-making style ($\chi^2 = 3.17$; $p = .205$), attitude towards AI ($\chi^2 = 1.68$; $p = .432$), or technostress ($\chi^2 = 2.15$; $p = .342$). Similarly, post-hoc comparisons did not reveal significant differences between groups.

No significant differences in observed changes emerged with respect to professional experience in school settings either. Specifically, results for DIFF_STILE ($\chi^2 = 0.172$; $p = .982$), DIFF_ATT_AI ($\chi^2 = 0.954$; $p = .812$), and DIFF_TECHNO ($\chi^2 = 3.104$; $p = .376$) indicate substantial homogeneity in the effects of the GenAI experience regardless of participants' employment condition. Effect sizes are also extremely small (ε^2 between .001 and .022).

No inferential comparisons by gender were conducted, given the strong asymmetry in the sample, which consists predominantly of women (96.5%).

Finally, differences in observed changes were analyzed in relation to the tool perceived as most useful for rubric improvement. Specifically, 54.9% of students indicated AI as the main tool, 42.4% saw integration of AI and exemplars as most useful, and small proportions indicated only the exemplar (2.1%) or neither (0.7%).

Group comparisons did not reveal statistically significant differences in change scores for learning style ($\chi^2 = 3.51$; $p = .319$), attitude towards AI ($\chi^2 = 3.23$; $p = .358$), or technostress ($\chi^2 = 3.77$; $p = .287$). Nor did significant differences emerge in perceptions of development of evaluative competences ($\chi^2 = 2.84$; $p = .416$) or of the tools' contribution to rubric improvement ($\chi^2 = 2.93$; $p = .402$). Effect sizes are again small (ε^2 between .019 and .026), suggesting that perceived benefits from the experience were largely shared across different profiles of tool use.

These results must nonetheless be interpreted with caution due to the strong asymmetry in group sizes for the "most useful tool" variable, especially in the "only exemplar" and "none" categories.

DISCUSSION

This study explored the extent to which integrating GenAI into an authentic assessment activity centred on rubric construction and revision is associated with changes in perceptions among pre-service university students. Results indicate statistically significant pre–post differences across all dimensions considered. However, given the single-group quasi-experimental design without a control group, these findings should be interpreted as evidence of co-occurrence rather than causal effects: observed changes cannot be unequivocally attributed to the GenAI intervention, as the influence of confounding factors — maturation, repeated testing, or general course exposure — cannot be ruled out.

CHANGES IN PERCEPTIONS AND DISPOSITIONS (RQ1)

Participation in the experience was associated with positive shifts across all three dimensions: reflective and criterion-based approaches (M: 4.78 to 5.06), attitude towards AI (M: 4.12 to 4.84), and perceived technostress, which decreased (M: 3.12 to 2.90). These associations are consistent with Kolb's (1984) experiential learning model, according to which learning develops through iterative cycles of concrete experience, reflection, and conceptual revision. The activity's structure - initial rubric construction, comparison with exemplars and AI-generated feedback, product revision, and self-assessment - may have supported such reflective processes, with students appearing to use GenAI as a cognitive scaffold rather than as an execution tool.

From an authentic assessment perspective (Wiggins, 1990; Ashford-Rowe et al., 2014), the task required students to exercise pedagogical judgement within a professionally meaningful context. Within this setting, GenAI appeared to operate less as a substitute for human judgement than as a dialogic resource that may support criterion interpretation, quality recognition, and evaluative reflection, consistent with the evaluative judgement framework (Lipnevich et al., 2014; 2023; Tai et al., 2018). The complementarity between exemplars and GenAI is particularly noteworthy: while exemplars provide stable reference points for calibration, GenAI may offer opportunities for iterative clarification and reformulation, potentially encouraging a more active negotiation of assessment criteria.

The increase in positive attitudes towards AI (large effect size: $r = -0.890$) is consistent with technology acceptance research, in which meaningful engagement with digital tools has been associated with stronger perceptions of usefulness (Davis, 1989; Venkatesh et al., 2003). Likewise, the reduction in technostress, although more modest ($r = 0.421$), aligns with previous studies suggesting that guided and pedagogically mediated interaction with digital technologies may contribute to lower perceived technological anxiety through greater familiarity, self-efficacy, and instructional support (Ragu-Nathan et al., 2008; Joo et al., 2016). Overall, these findings are consistent with the interpretation that educational mediation and task authenticity, rather than mere exposure to technology, may have played an important role. Nevertheless, this interpretation remains tentative given the absence of a control group and should not be interpreted as evidence of a causal effect.

CHANGES BETWEEN PRE-TEST AND POST-TEST: MAGNITUDE, STATISTICAL SIGNIFICANCE, AND EFFECT SIZE (RQ2)

Despite the overall positive pre–post trends, individual difference scores reveal considerable heterogeneity in participants' responses. The largest mean change was observed for attitude towards AI ($M = 0.72$; $SD = 0.75$), followed by reflective and criterion-based approaches ($M = 0.29$; $SD = 0.41$), whereas technostress showed a more modest mean reduction ($M = -0.22$; $SD = 0.57$). The wide range of individual scores indicates that some students experienced substantial improvements, whereas others showed limited or even negative changes.

Such variability is consistent with previous research suggesting that engagement with AI is shaped by individual characteristics including digital self-efficacy, prior experience, trust, attitudes, and perceived control (Symasek et al., 2025; Kelly et al., 2023). Accordingly, the present findings suggest that students exposed to the same instructional experience may engage with GenAI in different ways, highlighting the importance of designing AI-supported assessment activities that accommodate diverse learner profiles rather than assuming uniform responses.

Inferential analyses confirmed statistically significant pre–post differences across all three dimensions. Wilcoxon tests indicated significant increases in reflective and criterion-based approaches ($W = 1033$; $p < .001$) and attitudes towards AI ($W = 439$; $p < .001$), together with a significant reduction in technostress ($W = 5774$; $p < .001$). Effect sizes were large for attitudes towards AI ($r = -0.890$) and reflective approaches ($r = -0.725$), whereas the effect for technostress was moderate ($r = 0.421$).

The larger effects observed for attitudinal and reflective dimensions are consistent with technology acceptance research, in which meaningful engagement with digital technologies has been associated with more favourable attitudes and perceived usefulness (Davis, 1989; Venkatesh et al., 2003). By contrast, the smaller reduction in technostress is consistent with literature describing emotional responses to technology as relatively stable and influenced by broader factors such as uncertainty, perceived complexity, and loss of control (Ragu-Nathan et al., 2008; Joo et al., 2016; Floridi et al., 2018).

Beyond statistical significance, these findings suggest that changes in students' perceptions may be associated not only with exposure to GenAI itself but also with its integration into a structured learning activity involving comparison with exemplars, iterative revision, and self-assessment. This interpretation is consistent with the evaluative judgement perspective (Tai et al., 2018), although, given the absence of a control group, it cannot be determined whether similar changes would have occurred following an equivalent learning activity without GenAI.

These findings should therefore be interpreted cautiously in light of the exploratory pre–post design, the absence of a comparison group, the use of self-report measures, and the relatively low internal consistency of the Reflective and Criterion-Based Approaches Scale. In particular, the modest reliability of this measure suggests that findings related to this construct should be considered with caution, as measurement error may have reduced the precision and stability of the observed associations and pre–post differences.

ASSOCIATION BETWEEN INITIAL LEVELS AND CHANGES AFTER THE GENAI EXPERIENCE (RQ3)

Results indicate that students' initial levels across the three dimensions were significantly associated with the magnitude of the observed changes following the learning experience. Significant negative correlations were found between pre-test scores and corresponding change scores for reflective and criterion-based approaches ($\rho = -0.421$; $p < .001$), attitudes towards AI ($\rho = -0.663$; $p < .001$), and technostress/computer anxiety ($\rho = -0.355$; $p < .001$). Overall, students with less favourable baseline profiles—lower openness towards AI, lower levels of reflective engagement, or higher technological discomfort—tended to exhibit larger changes following the intervention.

The strongest association was observed for attitudes towards AI, although similar patterns emerged for technostress and reflective approaches. Across all three dimensions, students with less favourable baseline profiles showed greater changes than those with more favourable initial scores. These findings are consistent with previous research suggesting that meaningful engagement with AI-supported learning activities may be associated with more positive attitudes towards AI, lower technological anxiety, and greater confidence when pedagogical support is provided (Davis, 1989; Venkatesh et al., 2003; Ragu-Nathan et al., 2008; Joo et al., 2016).

The association observed for reflective and criterion-based approaches is also consistent with the interpretation that the learning activity may have provided opportunities for students with lower initial levels of reflective engagement to participate more actively in processes of comparison, revision, and criterion-based judgement, in line with the evaluative judgement perspective (Tai et al., 2018; Lipnevich et al., 2014; 2023.

Taken together, these findings suggest that students' initial dispositions may be an important factor in understanding how they engage with GenAI-supported assessment activities. Consistent with Symasek et al. (2025), individual differences appear to play a central role in AI acceptance and educational use. However, these associations should be interpreted cautiously, as correlations between baseline scores and change scores may partially reflect regression to the mean and the mathematical dependence between these variables rather than substantive relationships alone.

DIFFERENCES IN OBSERVED CHANGES IN RELATION TO INDIVIDUAL CHARACTERISTICS AND TOOLS USED (RQ4)

Comparative analyses did not reveal statistically significant differences in observed changes according to the individual characteristics examined or the tool perceived as most useful during rubric revision. Specifically, no significant differences emerged in changes in reflective and criterion-based approaches, attitudes towards AI, or technostress according to generational group or professional experience in school settings. Likewise, students who identified AI, exemplars, the combination of both, or neither as the most useful support did not differ significantly in either observed changes or perceptions of evaluative competence development and rubric improvement.

Overall, these findings suggest that the observed changes were broadly comparable across the groups considered. The absence of generational differences provides limited support for common assumptions that younger cohorts are inherently more predisposed to engaging with AI technologies, and instead is consistent with research suggesting that contextual and pedagogical factors may be more influential than generational membership alone (Symasek et al., 2025). Similarly, students with prior professional experience in school settings did not differ from those without such experience, indicating that the learning activity may have represented a relatively novel experience regardless of professional background.

The absence of differences according to the tool perceived as most useful is also noteworthy. Although many students identified GenAI as their primary source of support, while others valued exemplars or the combination of both, similar patterns of change were observed across groups. This finding suggests that the educational value of the experience may have been associated less with any single resource than with participation in a structured process involving comparison, revision, feedback, and reflection.

These findings should nevertheless be interpreted cautiously. The absence of statistically significant group differences does not imply that individual characteristics are unimportant. As shown in RQ4, students' initial dispositions were significantly associated with the magnitude of observed changes, suggesting that dispositional variables may be more informative than broad categorical characteristics such as age or professional experience. Furthermore, the relatively homogeneous sample - predominantly female Generation Z students - and the very small effect sizes observed in group comparisons may have limited the ability to detect meaningful between-group differences.

Future research should therefore examine these relationships using more heterogeneous samples and considering additional learner characteristics, including digital competence, prior familiarity with AI, technological self-efficacy, trust in AI, and self-regulation strategies.

CONCLUSIONS

This study explored the integration of GenAI into an authentic assessment and experiential learning activity centered on rubric construction, revision, and self-assessment by future primary school teachers. Overall, results seem to show that the GenAI experience is associated with positive changes in students' dispositions, with significant increases in reflective and criterion-based approaches and in attitude towards AI, accompanied by a reduction in perceived technostress/computer anxiety.

In line with the authentic assessment paradigm (Wiggins, 1990; Ashford-Rowe et al., 2014; Ajjawi et al., 2023), findings suggest that GenAI can play a pedagogically meaningful role when integrated into authentic, reflective, and intentionally designed activities. The experience appears to have fostered critical comparison, revision, and development of evaluative judgement (Tai et al., 2018), supporting more informed, criterion-oriented modes of engagement in assessment processes. Consistent with Kolb's (1984) experiential learning model, the combination of rubric construction, comparison with exemplars and AI-generated feedback, revision, and self-assessment seems to have activated cyclic processes of reflection, re-elaboration, and product improvement.

Findings also align with the perspective of GenAI as an agent-to-learn-with (Salinas-Navarro et al., 2024), according to which AI can support dialogic, metacognitive, and reflective processes rather than merely automating tasks. In this sense, GenAI appears to have functioned as a cognitive and evaluative scaffold (Jonassen, 1996; Kim & Baylor, 2006), supporting students in critical revision, criterion-based comparison, and self-regulation of learning.

The study further highlights marked individual variability in observed changes. In particular, students with less favorable initial attitudes towards AI or higher technostress tend to show more substantial changes after the experience. These findings are consistent with literature on technology acceptance and individual differences (Davis, 1989; Venkatesh et al., 2003; Symasek et al., 2025), which emphasizes that attitudes, self-efficacy, trust, and initial dispositions significantly influence interaction with AI technologies in education.

At the same time, effects appear relatively transversal with respect to some socio-demographic and professional characteristics. The absence of significant differences by generational group, professional experience in school settings, or perceived most useful tool suggests that the formative value of the experience may depend more on the quality of pedagogical design and activity structure than on users' demographic or professional characteristics alone (Laurillard, 2012; Mishra & Koehler, 2006).

Nonetheless, results must be interpreted with caution. The sample is relatively homogeneous, consisting mainly of female (96.5%) Generation Z (83.3%) students from a single university, limiting generalizability. Furthermore, the absence of a control group and the self-report nature of measures do not allow firm causal attribution of observed changes to GenAI use or direct assessment of actual improvement in students' evaluative competences. Additional caution is warranted regarding the reflective and criterion-based approaches scale, which showed lower internal consistency than the other dimensions.

Future research could include more heterogeneous samples, performance-based measures, and longitudinal designs, and further examine individual variables such as technological self-efficacy, trust in AI systems, familiarity with AI, and prior digital competences. It would also be useful to explore different modalities of GenAI integration in authentic assessment processes, analyzing how varying levels of pedagogical mediation influence development of evaluative judgement, reflective competences, and self-regulation.

Overall, this study suggests that GenAI can be a promising resource to support authentic, reflective, and participatory assessment practices in higher education, provided it is integrated into pedagogically designed environments oriented towards the critical development of students' evaluative judgement rather than mere automation of assessment processes.

REFERENCES

Agarwal, R., & Prasad, J. (1999). Are individual differences germane to the acceptance of new information technologies? *Decision Sciences, 30*(2), 361–391. https://doi.org/10.1111/j.1540-5915.1999.tb01614.x

Ajjawi, R., Tai, J., Huu Nghia, T. L., Boud, D., Johnson, L., & Patrick, C.-J. (2023). Aligning assessment with the needs of work-integrated learning: The challenge of authentic assessment. *Assessment & Evaluation in Higher Education, 48*(2), 149–160. https://doi.org/10.1080/02602938.2022.2048800

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education, 39*(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI, 7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bearman, M., Dawson, P., Ajjawi, R., Tai, J., & Boud, D. (Eds.). (2022). *Re-imagining university assessment in a digital world*. Springer. https://doi.org/10.1007/978-3-030-73521-8

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413. https://doi.org/10.1080/02602938.2015.1018133
Brookhart S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading.* ASCD, Alexandria
Chae, H.-C., Koh, C. E., & Jeon, M. (2020). Can cognitive style predict adoption of an emerging technology? A study of cognitive style and its influence on the perception of a new technology. *International Journal of Technology and Human Interaction, 16*(1), 18–35. https://doi.org/10.4018/IJTHI.2020010102
Chakraborty, I., Hu, P. J.-H., & Cui, D. (2008). Examining the effects of cognitive style in individuals' technology use decision making. *Decision Support Systems, 45*(2), 228–241. https://doi.org/10.1016/j.dss.2007.02.002
Chiu, T. K. F. (2023). The impact of generative AI (GenAI) on practices, policies, and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2023.2253861
Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly, 19*(2), 189–211. https://doi.org/10.2307/249688
Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice, 20*(3), Article 02. https://doi.org/10.53761/1.20.3.02
Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20*, Article 22. https://doi.org/10.1186/s41239-023-00392-8
Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008
Demir, T., & Çüm, S. (2026). Rubric-Based Assessment of Narrative Texts via Human-AI Collaboration: A Specialized GPT Model Approach Evaluación de Textos Narrativos Basada en Rúbricas con Colaboración Humano-IA. *Revista de Educación*, 411, 339-372.
Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer. https://doi.org/10.1007/978-3-030-30371-6
Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. Pew research center, 17(1), 1-7.
Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*, Article 57. https://doi.org/10.1186/s41239-023-00425-2
Faqih, K. M. S. (2011). Integrating perceived risk and trust with technology acceptance model: An empirical assessment of customers' acceptance of online shopping in Jordan. *2011 International Conference on Research and Innovation in Information Systems*, 1–5. IEEE. https://doi.org/10.1109/ICRIIS.2011.6125693
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5
Fry, H., Ketteridge, S., & Marshall, S. (Eds.). (2023). *A handbook for teaching and learning in higher education: Enhancing academic practice* (5th ed.). Routledge.
Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has aced the test of understanding in college economics: Now what? *American Economist, 68*(2), 233–245. https://doi.org/10.1177/05694345231169654
Ifenthaler, D., & Schumacher, C. (2023). Reciprocal issues of artificial and human intelligence in education. *Journal of Research on Technology in Education, 55*(1), 1–6. https://doi.org/10.1080/15391523.2022.2154511
Jia, F., Sun, D., & Looi, C. K. (2024). Artificial intelligence in science education (2013–2023): Research trends in ten years. *Journal of Science Education and Technology*, 33(1), 94-117.
Jonassen, D. H. (1996). *Computers in the classroom: Mindtools for critical thinking*. Prentice-Hall.
Joo, Y. J., Park, S., & Lim, E. (2016). Factors influencing mobile learning acceptance: Self-efficacy, self-directed learning, and technology acceptance. *Educational Technology & Society, 19*(3), 214–224.
Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kelly, S., McLaughlan, R., & Kaufman, J. (2023). A systematic review of AI acceptance: Emerging factors and trends. *Computers in Human Behavior, 148*, Article 107865. https://doi.org/10.1016/j.chb.2023.107865
Kim, Y., & Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development, 54*(6), 569–596. https://doi.org/10.1007/s11423-006-0637-3
Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60–70.
Koh, K. H. (2017). Authentic assessment. In G. W. Noblit (Ed.), *Oxford research encyclopedia of education*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190264093.013.22
Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.
Kolb, A. Y., & Kolb, D. A. (2018). Eight important things to know about the experiential learning cycle. *Australian Educational Leader, 40*(3), 8–14.
Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.
Lipnevich, A. A., Panadero, E., & Calistro, T. (2023). Unraveling the effects of rubrics and exemplars on student writing performance. *Journal of Experimental Psychology:* Applied, 29(1), 136.
Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, 42(4), 539-559.
Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81
Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x
Moon, J. A. (2004). *A handbook of reflective and experiential learning: Theory and practice*. RoutledgeFalmer.
Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422
Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002
Ragu-Nathan, T. S., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417–433. https://doi.org/10.1287/isre.1070.0165
Riding, R., & Rayner, S. (1998). *Cognitive styles and learning strategies*. David Fulton Publishers.
Rosen, L. D., & Weil, M. M. (1995). Computer anxiety: A cross-cultural comparison of university students in ten countries. *Computers in Human Behavior, 11*(1), 45–64. https://doi.org/10.1016/0747-5632(94)00021-9
Salinas-Navarro, D. E., Alvarado-Vanegas, B., & Romo-Bucheli, D. (2024). Generative AI for experiential learning and authentic assessment in higher education. *Computers & Education: Artificial Intelligence, 7*, Article 100254. https://doi.org/10.1016/j.caeai.2024.100254
Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., Sariyska, R., Stavrou, M., Becker, B., & Montag, C. (2022). Assessing the attitude towards artificial intelligence: Introduction of a short measure in German, Chinese, and English language. *KI – Künstliche Intelligenz, 35*(1), 109–118. https://doi.org/10.1007/s13218-020-00689-0
Steinert, Y., Mann, K., Anderson, B., Barnett, B. M., Centeno, A., Naismith, L., Prideaux, D., Spencer, J., Tullo, E., Viggiano, T., Ward, H., & Dolmans, D. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. *Medical Teacher, 38*(8), 769–786. https://doi.org/10.1080/0142159X.2016.1181851
Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature, 614*(7947), 214–216. https://doi.org/10.1038/d41586-023-00340-6
Svendsen, G. B., Johnsen, J.-A. K., Almås-Sørensen, L., & Vittersø, J. (2013). Personality and technology acceptance: The influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology, 32*(4), 323–334. https://doi.org/10.1080/0144929X.2011.553740
Symasek, L., Yeazitzis, T., Weger, K., & Mesmer, B. (2025). Recent developments in individual difference research to inform the adoption of AI technology. *Systems, 13*(3), 156. https://doi.org/10.3390/systems13030156

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Thatcher, J. B., & Perrewé, P. L. (2002). An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. *MIS Quarterly, 26*(4), 381–396. https://doi.org/10.2307/4132314

Varma, S., & Marler, J. H. (2013). The dual nature of prior computer experience: More is not necessarily better for technology acceptance. *Computers in Human Behavior, 29*(4), 1475–1482. https://doi.org/10.1016/j.chb.2013.01.024

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation, 2*(2). https://doi.org/10.7275/ffb1-mm19

Wu, I.-L., & Chen, J.-L. (2005). An extension of trust and TAM model with TPB in the initial adoption of online tax: An empirical study. *International Journal of Human–Computer Studies, 62*(6), 784–808. https://doi.org/10.1016/j.ijhcs.2005.03.003

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*, Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, X., Gao, Y., Yan, X., de Pablos, P. O., Sun, Y., & Cao, X. (2017). From e-learning to social-learning: Mapping development of studies on social-media-supported knowledge management. *Computers in Human Behavior, 72*, 140–148. https://doi.org/10.1016/j.chb.2017.02.038

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

APPENDIX A. Questionnaire: Perceptions, Attitudes, and Individual Differences in the Use of Artificial Intelligence for Formative Assessment

Section 1 – Alphanumeric Identification Code
Participants were asked to create an anonymous alphanumeric identification code.
Section 2 – Personal Information
(Group-analysis data only; non-identifiable information)
Age: __________
Gender:
- Woman
- Man
- Other
- Prefer not to say

Do you belong to one of the following generational groups?
- Gen Z (1997–2012)
- Millennial (1981–1996)
- Other / I do not know

Have you previously used assessment rubrics?
- Yes
- No
- I do not remember

Are you currently working in a school setting?
- Yes, in preschool education
- Yes, in primary education
- Yes, in other school levels
- No

Section 3 – Learning and Decision-Making Style
Response scale (self-anchoring 1–6):[Text Wrapping Break] 1 = Not at all true for me | 6 = Completely true for me

Item	Statement
1	I like learning by observing examples and reference models.
2	I tend to analyze educational situations systemically, looking for connections among elements.
3	I prefer making assessment decisions based on clear and structured criteria.
4	I feel more confident when I have a guide or framework to follow.
5	I tend to reflect on learning processes and on my own assessment criteria.

Section 4 – Attitudes Toward the Use of AI in Educational Contexts
Response scale (self-anchoring 1–6): [Text Wrapping Break] 1 = Not at all true for me | 6 = Completely true for me

Item	Statement
6	I feel comfortable using Artificial Intelligence tools in educational contexts.
7	I believe that AI can help me improve my assessment-related competences.
8	I find it interesting to explore the potential of AI for instructional design.
9	I feel motivated to use AI if it can provide personalized feedback.

10	I think that the use of AI can make assessment more objective and transparent.

Section 5 – Technostress and Technological Anxiety
Response scale (self-anchoring 1–6): [Text Wrapping Break] 1 = Not at all true for me | 6 = Completely true for me

Item	**Statement**
11	When I use new technologies, I feel anxious about making mistakes.
12	I feel stressed when I have to manage unfamiliar digital tools.
13	I am afraid of losing data or not being able to use AI-based tools correctly.
14	I would prefer to avoid the use of complex technologies in assessment.
15	Despite initial difficulties, I am able to adapt quickly to new digital tools.