*education sciences*

MDPI

*Article*

# Integrating Universal Design for Learning, Challenge-Based Learning, and Generative AI in Teacher Education: A Participatory Action Research Study

**Arasay Padrón Alvarez [1,*] and Odiel Estrada-Molina [2]**

[1] Faculty of Education and Humanities, Universidad Internacional de La Rioja, 26006 Logroño, Spain
[2] Department of Pedagogy, Universidad de Valladolid, 42004 Soria, Spain; odiel.estrada@uva.es
* Correspondence: arasay.padron@unir.net

**Abstract**

Teacher education incorporates generative artificial intelligence (GenAI) without developed inclusive pedagogical frameworks. This study contextually evaluates the Integrated Inclusive AI Challenge Model (MIIA-R). A concurrent mixed-methods design based on Participatory Action Research comprised three phases in a postgraduate programme and examined participation, engagement, and pedagogical creativity in designing inclusive GenAI-mediated proposals. The sample included 83 students, 80 proposals, and 178 learning artefacts. An independent rater blinded to the hypotheses re-evaluated a stratified subsample (ICC = 0.84; weighted $\kappa = 0.74$–$0.78$). Although the overall rubric score exceeded the functional threshold (M = 3.29), this reflected proposal feasibility (M = 4.18) and curricular coherence (M = 3.62), while UDL integration (M = 2.88) and pedagogical use of GenAI (M = 2.39) remained below it; therefore, Hypothesis 1 received partial and limited support. Pedagogical creativity followed a non-linear trajectory and correlated with GenAI tool diversity in the final phase ($\rho = 0.527$, $p < 0.001$). Profile differences were significant, $H(3) = 18.34$, $p < 0.001$, $\varepsilon^2 = 0.37$; Dunn tests showed that multimodal integrated use differed from generative-only and no or minimal use, but not from guided instrumental use. The MIIA-R supported feasible proposals but not consistently deep inclusive, AI-mediated pedagogical integration.

**Keywords:** Universal Design for Learning; challenge-based learning; generative artificial intelligence; creativity; teacher education

Check for updates

Academic Editors: Sergio Miranda and Frederico Corni

Received: 24 July 2026
Revised: 24 August 2026
Accepted: 25 August 2026
Published: 5 September 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

Higher education is undergoing a profound transformation driven by the rapid expansion of Generative Artificial Intelligence (GenAI) and the increasing demand for educational models capable of combining technological innovation with inclusion, accessibility, and pedagogical quality (García, 2021; Crompton & Burke, 2023). This transformation is particularly significant in teacher education, where future educators are expected not only to use GenAI effectively but also to integrate it critically, ethically, and pedagogically into inclusive learning environments. However, many higher education institutions are still adapting to this technological shift, creating a mismatch between the availability of increasingly sophisticated AI tools and the pedagogical frameworks required for their meaningful educational use (Holmes et al., 2019; Echeverria & Otero, 2025).

Among the pedagogical approaches currently informing educational innovation, Universal Design for Learning (UDL), Challenge-Based Learning (CBL), and GenAI each

address complementary dimensions of contemporary teaching and learning. UDL (CAST, 2024) conceptualises learner variability as the norm rather than the exception and provides a systematic framework for reducing learning barriers through multiple means of engagement, representation, and action and expression. CBL structures learning around authentic challenges that promote collaboration, critical thinking, creativity, and professional decision-making in real educational contexts (Johnson et al., 2009; Moreira-Alcivar, 2025). Meanwhile, GenAI has introduced unprecedented opportunities for generating learning resources, providing adaptive feedback, supporting instructional design, and fostering personalised learning, while simultaneously raising important ethical, pedagogical, and regulatory concerns (Zawacki-Richter et al., 2019; Xavier & Dutra, 2026). The technological pedagogical content knowledge (TPACK) framework also points in this direction by stressing that the meaningful integration of digital technologies requires a coherent articulation of content, pedagogy, and technology, which is especially relevant in teacher education programmes that incorporate challenge-based approaches and generative AI tools (Koehler et al., 2013).

Despite these advances, current research continues to examine these approaches largely in isolation. Studies on GenAI have predominantly focused on technological capabilities, automation, or personalised learning, often paying limited attention to the pedagogical principles that should guide its educational use (López et al., 2026). Research on UDL has consistently demonstrated its contribution to accessibility and participation, whereas CBL has shown positive effects on engagement, motivation, and competence development (Moral & Moreno-Tallón, 2025; Bastidas et al., 2025). However, these contributions have generally evolved along parallel lines rather than through integrated pedagogical frameworks capable of simultaneously addressing inclusion, authentic learning, and AI-mediated educational design.

Consequently, an important gap remains in understanding how these complementary perspectives can be articulated within a coherent pedagogical architecture. Existing studies rarely integrate UDL, CBL, and GenAI simultaneously, and even fewer provide empirical evidence derived from authentic online teacher education contexts. As a result, there is still limited knowledge regarding how inclusive pedagogical design, authentic challenge-based learning, and the pedagogically grounded use of GenAI interact to foster student participation, engagement, and pedagogical creativity (D. Dang et al., 2026).

Addressing this gap requires more than simply combining existing approaches. The educational challenge is not the absence of UDL, CBL, or GenAI individually, but rather the lack of pedagogical models capable of establishing functional relationships among them. In many current implementations, these approaches coexist without forming an integrated instructional design in which each component performs a complementary and pedagogically meaningful role. This limitation reduces the capacity of higher education programmes to exploit the educational potential of GenAI while maintaining coherent principles of inclusion and authentic learning.

To respond to this challenge, the present study proposes the Integrated Inclusive AI Challenge Model (MIIA-R), conceived as a pedagogical architecture that functionally integrates UDL, CBL, and GenAI within online teacher education. Rather than representing an additive combination of existing approaches, the model establishes an operational relationship in which UDL provides the inclusive pedagogical principles, CBL structures authentic learning processes, and GenAI functions as a pedagogical mediator that supports exploration, instructional design, adaptation, and feedback throughout the learning process. The model was developed and refined through iterative Participatory Action Research (PAR) cycles, enabling both its implementation and its context-sensitive empirical evaluation in an authentic educational setting.

Recent research suggests that the quality of inclusive learning design depends not only on the adoption of innovative pedagogical approaches but also on the way these approaches interact to support participation, engagement, and creativity (Gómez & Cruz, 2025; Veiga et al., 2026). These three dimensions are increasingly recognised as complementary indicators of meaningful learning in higher education, particularly within digitally mediated environments.

Student engagement, for example, extends beyond behavioural participation to encompass cognitive, affective–volitional, and agentic processes that influence how learners interpret challenges, regulate their learning, and construct knowledge in authentic educational contexts (Reeve, 2024). Similarly, pedagogical creativity involves more than the production of original artefacts; it reflects learners' capacity to design educational responses that are contextually relevant, inclusive, and pedagogically justified. Participation, in turn, represents the extent to which learning environments reduce barriers and create equitable opportunities for meaningful involvement.

From this perspective, UDL, CBL, and GenAI can be understood as addressing complementary dimensions of the same educational challenge rather than as independent pedagogical innovations. UDL contributes the principles required to ensure accessibility, flexibility, and meaningful participation. CBL structures learning around authentic professional challenges that promote sustained engagement, collaboration, and knowledge transfer. GenAI extends these processes by supporting exploration, instructional design, adaptive feedback, and the creative refinement of educational solutions. Their educational value therefore lies not simply in their individual contributions but in the functional relationships established among them.

This functional integration constitutes the theoretical foundation of the Integrated Inclusive AI Challenge Model (MIIA-R). Rather than conceptualising UDL, CBL, and GenAI as parallel components, the model assumes that participation, engagement, and pedagogical creativity emerge from the interaction among inclusive design, authentic challenge-based learning, and pedagogically grounded technological mediation. Consequently, the present study examines whether this integrated pedagogical architecture contributes to improving these dimensions within an online postgraduate teacher education programme.

Accordingly, the study is guided by the following research question, objective, and hypotheses.

**Research question.** *How can student participation, engagement, and pedagogical creativity be enhanced in the design of inclusive teaching and learning proposals mediated by GenAI within an online postgraduate programme?*

**Aim.** To design, implement, and empirically evaluate the MIIA-R as a pedagogical framework for developing inclusive teaching and learning proposals mediated by GenAI in an online postgraduate programme while examining its contribution to student participation, engagement, and pedagogical creativity.

**H1.** *Implementation of the MIIA-R will be associated with sustained student participation, active engagement throughout the learning process, and the development of pedagogically robust teaching and learning proposals whose overall quality exceeds the predefined functional threshold ($M \geq 3.0$).*

**H2.** *Greater diversity in the pedagogically meaningful use of GenAI tools will be positively associated with higher levels of pedagogical creativity in the resulting teaching and learning proposals.*

Taken together, these observations suggest that the current challenge in higher education is not the lack of pedagogical or technological approaches, but the absence of an integrated instructional architecture capable of orchestrating them coherently within authentic learning contexts. To address this challenge, the Integrated Inclusive AI Challenge

Model (MIIA-R) proposes a pedagogical architecture that functionally integrates Universal Design for Learning (UDL), Challenge-Based Learning (CBL), and Generative Artificial Intelligence (GenAI). Within this architecture, UDL provides the principles of inclusion and accessibility, CBL structures learning around authentic practice-oriented challenges, and GenAI functions as a pedagogical mediator that supports exploration, instructional design, adaptation, and feedback. Figure 1 illustrates the conceptual architecture of the model, while its methodological development and operational implementation are described in Sections 2.3 and 2.4. A detailed description of the model's operational framework is provided in Supplementary Materials (Section S6).

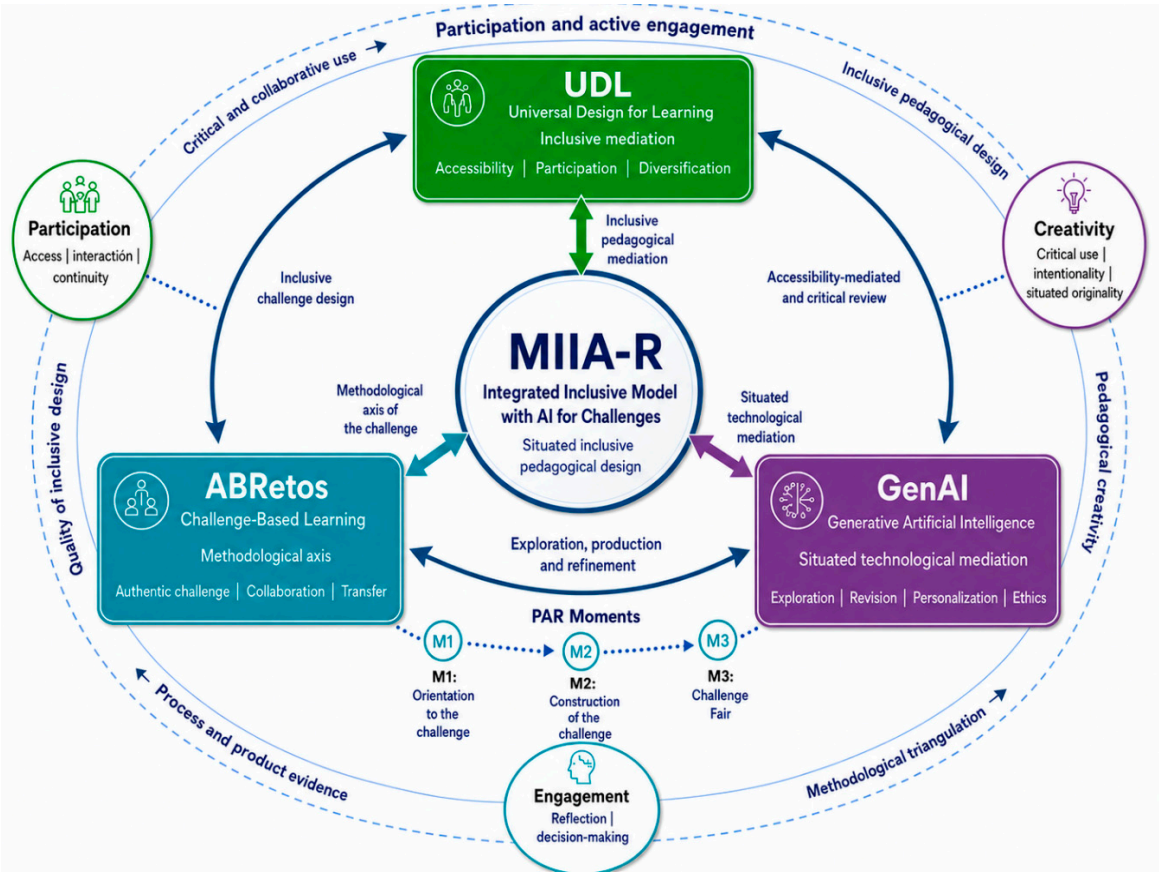



**Figure 1.** Architecture of the MIIA-R Model.

The originality of the MIIA-R lies not in combining UDL, CBL, and GenAI per se, but in establishing a functionally integrated pedagogical architecture in which these approaches perform complementary and interdependent roles. Functional integration is understood as an operational relationship in which each component fulfils a distinct educational function: UDL provides the principles of accessibility, participation, and flexible learning design; CBL organises learning through authentic and contextually relevant challenges; and GenAI operates as a pedagogical mediator embedded within the instructional design, supporting exploration, content creation, adaptation, and feedback. From this perspective, participation, engagement, and pedagogical creativity are conceptualised as emergent educational outcomes arising from the interaction among inclusive design, authentic learning activities, and technology-mediated pedagogical support, consistent with systemic and ecological perspectives on learning.

Accordingly, the MIIA-R is proposed as an adaptable pedagogical architecture for teacher education programmes seeking to integrate GenAI while maintaining a strong

commitment to inclusive pedagogy and authentic problem-solving. Rather than proposing a new theory of learning, its contribution lies in the conceptual integration and operational articulation of established educational approaches. The following sections empirically evaluate the model within an authentic online teacher education context.

Accordingly, this study makes three complementary contributions. Conceptually, it introduces a pedagogical architecture based on the functional integration of UDL, CBL, and GenAI. Methodologically, it develops and empirically evaluates the MIIA-R through iterative Participatory Action Research (PAR) supported by a mixed-methods design, providing triangulated quantitative and qualitative evidence of its feasibility and educational relevance within an authentic online teacher education context. Practically, it offers an evidence-informed framework together with practical implementation guidelines to support the design of inclusive AI-mediated learning environments in teacher education.

## 2. Materials and Methods

### *2.1. Research Approach and Design*

The methodological approach was selected to address the central purpose of the study: the iterative development, refinement, and contextual evaluation of the Integrated Inclusive AI Challenge Model (MIIA-R) within an authentic online teacher education environment. Rather than estimating causal effects under controlled conditions, the study sought to examine how the proposed pedagogical architecture functioned in practice while generating evidence to inform its continuous improvement. For this reason, Participatory Action Research (PAR) provided an appropriate methodological framework, as it supports cyclical processes of planning, action, observation, and reflection aimed at improving educational practice in context (Herr & Anderson, 2015; Kemmis & McTaggart, 2005).

To capture the complexity of this educational process, the study adopted a concurrent mixed-methods design integrating quantitative and qualitative evidence collected throughout the intervention. Quantitative data provided indicators of participation, pedagogical quality, creativity, and model performance, whereas qualitative evidence contributed insights into participants' engagement, decision-making processes, collaborative practices, and professional learning experiences. The concurrent integration of both forms of evidence enabled methodological triangulation and strengthened the ecological validity and contextual credibility of the findings (Creswell & Creswell, 2018; Creswell & Clark, 2017).

The study was conducted within a socio-critical transformative paradigm, which views educational research as a process of knowledge generation closely linked to the improvement of professional practice. Consistent with this perspective, participants acted as collaborators throughout successive PAR cycles, contributing to the refinement of the pedagogical model while simultaneously generating evidence regarding its implementation. Methodological rigour was enhanced through triangulation across data sources, instruments, and the three iterative PAR cycles (Denzin, 2017; Contreras-Villalobos et al., 2024).

In line with this socio-critical orientation, the study also engages explicitly with the theoretical tradition of adult education and, in particular, with transformative learning theory (Mezirow, 1991; Taylor & Cranton, 2012). Postgraduate teacher education constitutes an adult learning context in which Challenge-Based Learning, the backbone of the MIIA-R, draws on the principles of andragogy and experiential learning (Knowles et al., 2015; Kolb, 2015): authentic challenges activate participants' prior professional experience as a learning resource and promote critical reflection on their pedagogical assumptions and frames of reference. From this perspective, the reflection documented in the reflective journals is not a merely procedural record of activity, but functions as a mechanism of critical reflection aimed at premise revision, collaborative dialogue, and the transformation of professional practice (Mezirow, 1991). This grounding prevents teacher education from being reduced to

a content-centred procedural logic and situates it around the learner as the subject of his or her own transformation, consistent with the human and personal approach to technology use in higher education that this journal promotes (Benade, 2015; Li & Walsh, 2023; van Rensburg & Matahela, 2025).

### 2.2. Context and Participants

The study was conducted within a compulsory module of a fully online Master's programme in Education. The module focuses on the design of inclusive teaching and learning proposals and integrates UDL, CBL, and GenAI both as curricular content and as pedagogical strategies. This context was particularly appropriate for evaluating the MIIA-R because participants experienced the same pedagogical approach that they were subsequently expected to apply in their own professional practice.

Participants were recruited through purposive sampling from students enrolled in the module during the implementation of the intervention. Participation in the study was voluntary and anonymous, and all participants provided informed consent for the use of their data for research purposes.

A total of 83 students participated in the initial diagnostic phase. At the conclusion of the intervention, 80 valid teaching and learning proposals were available for rubric-based assessment. Because several proposals were developed collaboratively, the analytical unit for rubric-based assessment was the teaching and learning proposal rather than the individual participant. Across the three iterative PAR cycles, a total of 178 interim and final learning artefacts were analysed.

Participants had a mean age of 34.2 years ($SD$ = 6.8); 72.3% were women, 56.6% resided in Spain, 97.6% were professionally active, and 72.3% were practicing teachers. Table 1 summarises the sociodemographic characteristics of the participants. The intervention was implemented during the second half of the academic module (weeks 4–12), between 15 April 2026 and 8 June 2026, following the final favourable assessment issued by the Research Ethics Committee on 14 April 2026. The definitive ethics certificate was signed on 21 April 2026. The initial diagnostic survey was administered after ethical approval, between 15 and 17 April 2026.

**Table 1.** Sociodemographic Profile of the Participants ($n$ = 83).

| Variable | Category | *n* | % |
|---|---|---|---|
| Sex | Female | 60 | 72.3 |
| | Man | 23 | 27.7 |
| Country/Region of Origin | Spain | 47 | 56.6 |
| | Latin America | 31 | 37.3 |
| | Other | 5 | 6.0 |
| Employment Status | Employed | 81 | 97.6 |
| | Unemployed | 2 | 2.4 |
| Professional Role | Practising teacher | 60 | 72.3 |
| | Other educational professional | 21 | 25.3 |
| Age | Mean ($SD$) | 34.2 (6.8) | — |

Note. $SD$ = standard deviation. Percentages may not total 100% due to rounding.

### 2.3. Operationalisation of the MIIA-R Model and Design of the Three PAR Phases

The intervention was organised into three interconnected iterative PAR phases, each building systematically upon the previous one. Phases 1 and 2 comprised preparatory learning activities designed to progressively develop the pedagogical and technological competences required for the Final Challenge. Activity 1 introduced the principles of

UDL through active learning methodologies, whereas Activity 2 focused on the ethical and pedagogically meaningful use of GenAI to support educational design. Following this instructional scaffolding, participants completed the Final Challenge during Phase 3, integrating the knowledge and competences developed throughout the preceding activities (Figures 1 and 2).

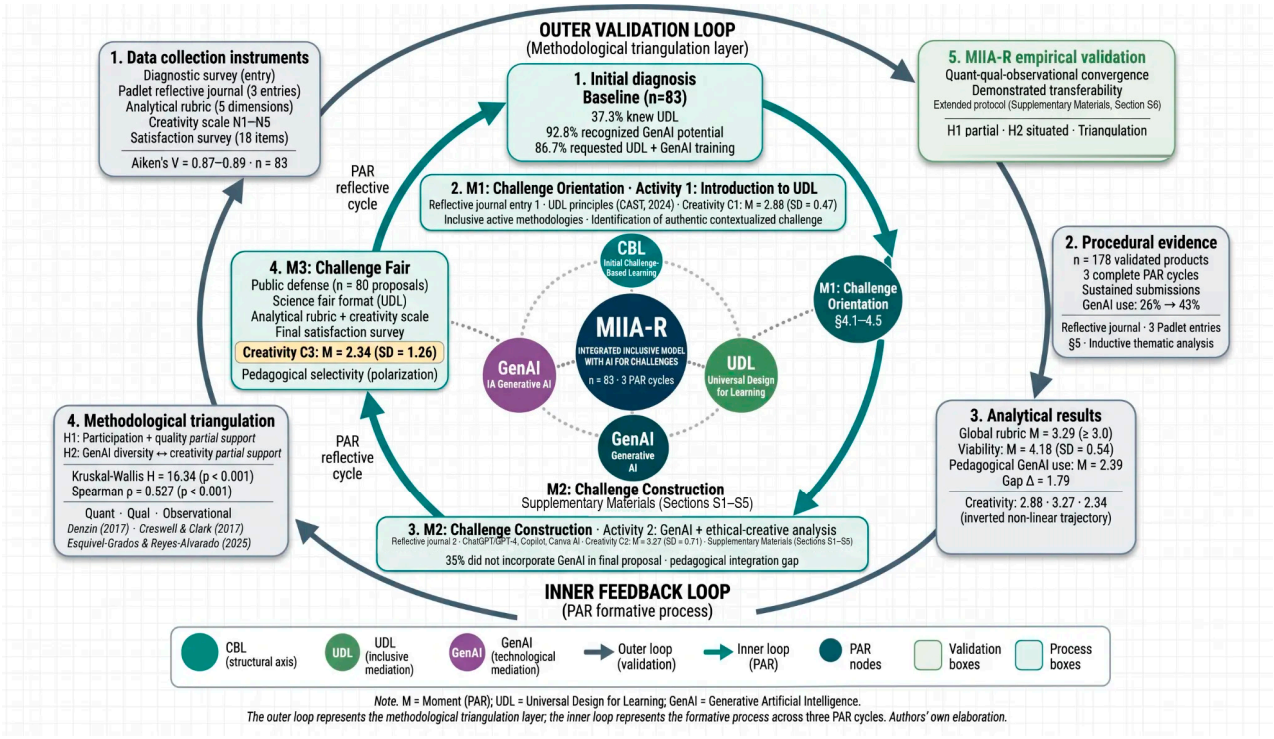



**Figure 2.** Learning Process and Implementation of the MIIA-R Across Three PAR Phases. References cited in the figure: CAST (2024), Denzin (2017), Creswell and Clark (2017), and Esquivel-Grados and Reyes-Alvarado (2025). Figure 1 presents the conceptual architecture of the MIIA-R, illustrating how CBL provides the organisational structure of the learning process, UDL ensures inclusive and accessible instructional design, and GenAI functions as a pedagogical mediator across the three iterative PAR phases. A detailed operational representation of the model, including its core components, implementation procedures, sources of evidence, and alignment with the PAR cycles, is provided in Supplementary Materials (Section S6).

### *2.4. Learning Sequence and Implementation Across the Three PAR Phases*

The MIIA-R was implemented through a structured learning sequence organised into three iterative PAR phases: Challenge Orientation, Challenge Development, and Challenge Showcase. Within this sequence, CBL provided the overall pedagogical structure by organising learning around the identification, design, and presentation of an instructional response to an authentic professional challenge. UDL ensured inclusive participation through accessible learning pathways and multiple forms of support, while GenAI functioned as a pedagogical mediator that facilitated exploration, instructional design, revision, and refinement throughout the intervention.

Each PAR phase incorporated specific learning activities, monitoring instruments, and formative feedback mechanisms aligned with the objectives of the study. This sequential design enabled the progressive development of participants' pedagogical proposals while simultaneously generating multiple sources of evidence for the contextual evaluation of the MIIA-R. Accordingly, the model was examined not only as a conceptual framework but also as an observable pedagogical process implemented in an authentic online teacher education context.

Figure 2 summarises the overall learning sequence and its relationship to the contextual evaluation of the MIIA-R. Table 2 provides an operational overview of each PAR phase, including the corresponding learning activities, research instruments, and expected outputs. Detailed descriptions of the intervention sequence, implementation procedures, research instruments, and learning activities are presented in Supplementary Materials (Sections S1–S6).

**Table 2.** Operational Sequence of the MIIA-R Model Across the Three PAR Phases.

| PAR Phase | Core Learning Activity | Key Instruments | Expected Outcome |
| --- | --- | --- | --- |
| Phase 1: Challenge Orientation | Initial diagnostic assessment + Activity 1: UDL-based learning sequences using active learning methodologies | Diagnostic survey; D1–D2 rubric; Padlet 1 | Baseline profile of the cohort; understanding of the challenge; initial inclusive instructional design |
| Phase 2: Challenge Development | Activity 2 + prototype development: integration of GenAI and ethical–creative analysis | Full analytical rubric; creativity scale; Padlet 2 | Proposal under development; guided use of GenAI; pedagogical refinement |
| Phase 3: Challenge Showcase | Final Challenge + assessment: public presentation and peer review | Overall analytical rubric; creativity scale; final survey; Padlet 3 | Validated final teaching and learning proposal; metacognitive reflection; evidence of transfer |

### *2.5. Data Collection Techniques and Instruments for Assessing Participation, Engagement, Creativity, and the Quality of Teaching and Learning Proposals*

Five complementary instruments were employed to characterise the baseline context, monitor the implementation process, and evaluate the outcomes of the intervention. Collectively, these instruments generated convergent quantitative and qualitative evidence regarding participation, engagement, pedagogical creativity, and the quality of the teaching and learning proposals. Their combined use enabled methodological triangulation across different stages of the intervention and multiple sources of evidence.

Initial diagnostic survey. Administered prior to the intervention, this survey identified participants' prior knowledge of UDL and GenAI, previous experience with these approaches, and perceived learning needs. It served exclusively to characterise the educational context and establish a baseline profile; it was not designed as a pretest for statistical comparison with post-intervention measures.

Group reflective journal (Padlet). Participants completed three structured reflective entries, one during each PAR phase. These journals provided qualitative evidence of participation, engagement, pedagogical decision-making, collaborative learning processes, and professional reflection. The complete protocol is available in Supplementary Materials (Section S2, Table S3).

Creativity Scale for GenAI Use. Pedagogical creativity was assessed using a five-level observational scale (N1–N5) developed specifically for this study and applied to the learning artefacts produced throughout the intervention. As an ad hoc observational instrument, its purpose was to classify observable levels of pedagogical creativity rather than to provide a standardised psychometric measure. Accordingly, the resulting scores were interpreted as context-specific evaluative evidence. The complete instrument is presented in Supplementary Materials (Section S1; Table S1 and Figure S1).

Operational definition of the construct. In this study, pedagogical creativity is defined as the capacity to generate instructional proposals that are simultaneously original, novel, relevant to the curricular context, and pedagogically inclusive, that is, proposals that

broaden students' options for participation, collaboration, representation, and expression. It is not defined as technical mastery of tools or as the accumulation of digital resources.

Distinction between technical complexity and pedagogical innovation. The use of several GenAI tools does not automatically equate to pedagogical creativity. Technical complexity refers to the number and sophistication of the tools or functions employed, whereas pedagogical innovation refers to the transformation of the instructional proposal: which learning barriers are removed, which inclusive decisions are justified, how students collaborate, and what added value GenAI mediation provides in comparison with a conventional solution. Accordingly, a technically sophisticated but pedagogically conventional artefact is placed at the lower levels of the scale, whereas a technologically simple artefact that offers an original and contextually relevant inclusive solution may reach the higher levels.

Justification of the boundaries between levels. The N1–N5 cut-off points were established through three cumulatively ordered criteria: (i) the presence or absence of intentional GenAI mediation; (ii) the degree of pedagogical transformation of the task (reproduction, adaptation, or redesign); and (iii) the explicit presence of inclusive decisions justified in accordance with UDL. Accordingly, N1 corresponds to absent or merely reproductive use; N2, to content generation without pedagogical adaptation; N3, to identifiable pedagogical adaptation without explicit inclusive justification; N4, to pedagogical redesign with justified inclusive decisions; and N5, to original pedagogical redesign with sustained, collaborative, and transferable inclusive integration. The decisive transitions occur between N2 and N3, marking the shift from producing content to transforming the proposal, and between N3 and N4, marking the shift from creating and adapting to justifying inclusion.

Illustrative examples. Anonymised examples of participant artefacts across the empirically observed levels of the scale, together with the criteria for the unobserved N5 level and the rationale for the boundaries between levels, are provided in Supplementary Materials (Section S1; Table S2).

Analytical Rubric for Teaching and Learning Proposals. The final teaching and learning proposals were evaluated across five dimensions: curricular coherence, integration of UDL, pedagogical use of GenAI, collaboration, and feasibility. Each dimension was rated on a five-point scale (1–5), with M = 3.0 established as the minimum functional quality threshold. The complete rubric is available in Supplementary Materials (Section S3; Tables S4–S9 and Figure S2).

Final satisfaction survey. At the conclusion of the intervention, participants completed an 18-item five-point Likert questionnaire, three open-ended questions, and one Net Promoter Score (NPS) item. This instrument provided complementary evidence regarding perceived engagement, professional transfer, and participants' evaluation of the MIIA-R. Internal consistency was satisfactory (Cronbach's $\alpha = 0.87$; McDonald's $\omega = 0.89$, 95% CI [0.84, 0.91]), while expert judgement supported adequate content validity (Aiken's $V = 0.80$–$0.91$).

Evidence supporting the validity and reliability of the research instruments was obtained through complementary procedures. Content validity of the newly developed instruments was examined through expert judgement using Aiken's $V$, yielding satisfactory values for both the analytical rubric and the final satisfaction survey. Internal consistency was estimated for the satisfaction survey using Cronbach's alpha and McDonald's omega. Internal consistency was not calculated for the analytical rubric because it was designed to assess multiple, conceptually distinct dimensions of pedagogical quality rather than a single latent construct. Given that the initial rubric scores were assigned by members of the research team, an independent rater subsequently re-evaluated a stratified subsample, and inter-rater reliability was estimated using ICC and weighted Cohen's $\kappa$. Future applications should extend independent double assessment to the complete corpus of

evidence. Detailed information on the development and content validation of the research instruments is provided in Supplementary Materials (Section S5), while the independent second-assessment procedure and its complete inter-rater agreement results are reported in Sections 2.5.3 and 3.3.2, whereas the operational framework and contextual evaluation of the MIIA-R are described in Supplementary Materials (Section S6).

The functional threshold of $M \geq 3.0$ was established a priori, before data collection, and corresponds to the midpoint of the five-point Likert scale of the analytical rubric, operationally defined as the minimum level of acceptable pedagogical quality in each dimension. This criterion follows established practice in the validation of rubric-based educational instruments, where the scale midpoint represents the expected performance standard for considering a proposal functionally adequate (Aiken, 1985). The threshold was fixed by the research team prior to implementation and was not adjusted post hoc on the basis of the results obtained.

#### 2.5.1. Verification of Statistical Assumptions

The inferential analyses were conducted using non-parametric procedures. This analytical strategy was selected primarily because the principal outcome measures were derived from ordinal instruments, including the five-point analytical rubric and the five-level Creativity Scale for GenAI Use (N1–N5). In addition, the study sought to identify patterns of association and between-group differences within an applied educational context rather than to estimate population parameters under interval-level assumptions (Field, 2018; Pallant, 2020).

To support this analytical decision, an exploratory assessment of the underlying statistical assumptions was undertaken following established recommendations for applied educational research (Field, 2018; Pallant, 2020). The distribution of quantitative variables and composite scores derived from the ordinal instruments was examined using the Shapiro–Wilk test, recommended for small to moderate sample sizes, and complemented by the Kolmogorov–Smirnov test with Lilliefors correction. Homogeneity of variance across GenAI use profiles was assessed using Levene's median-based test because of its robustness to departures from normality.

The exploratory analyses identified statistically significant departures from normality in several key variables, particularly those related to the pedagogical use of GenAI, pedagogical creativity during the final PAR phase, and the diversity of GenAI tools employed. Together with the ordinal nature of the measurement instruments and the observed heterogeneity of variances across participant profiles, these findings provided additional support for the use of non-parametric procedures.

Accordingly, Spearman's rank-order correlation coefficient ($\rho$) was used to examine the association between the diversity of GenAI tool use and pedagogical creativity, whereas the Kruskal–Wallis H test was employed to compare pedagogical creativity scores across GenAI use profiles. Effect sizes ($\varepsilon^2$) were calculated for the Kruskal–Wallis analyses to facilitate the interpretation of practical significance.

Descriptive statistics were interpreted together with measures of dispersion, 95% confidence intervals, and the pedagogical relevance of the findings, without assuming interval-scale properties for the mean scores. Quantitative analyses were performed using IBM SPSS Statistics 29 with a significance level of $\alpha = 0.05$. Qualitative analyses were conducted through systematic manual coding supported by traceability matrices developed in Microsoft Excel. Table 3 summarises the assessment of statistical assumptions and the analytical decisions adopted for each inferential procedure.

**Table 3.** Verification of Statistical Assumptions and Analytical Decisions.

| Analysis | Variable(s) | Scale/n | Exploratory Assessment | Analytical Decision |
| --- | --- | --- | --- | --- |
| Overall quality of teaching and learning proposals (*H1*) | Overall rubric score | Aggregated ordinal scale (1–5); $n = 80$ | Normality was explored using the Shapiro–Wilk test and the Kolmogorov–Smirnov test with Lilliefors correction. The aggregated ordinal scores showed departures from normality and were therefore interpreted as ordinal data. | Descriptive analysis using *M*, *SD*, and 95% confidence intervals; complementary one-sample Wilcoxon signed-rank test against the predefined functional threshold ($M = 3.0$); findings interpreted with appropriate caution given the ordinal nature of the data. |
| Rubric dimensions | D2: UDL integration; D3: pedagogical use of GenAI; D5: feasibility | Ordinal scale (1–5); $n = 80$ | Skewed distributions and ceiling or floor effects were observed in some rubric dimensions. | Descriptive statistics and 95% confidence intervals; cautious interpretation of each dimension. |
| Creativity across PAR phases | Creativity level (N1–N5) | Phase 1: $n = 39$; Phase 2: $n = 52$; Phase 3: $n = 45$ | The creativity scale was ordinal and showed varying levels of dispersion across the three PAR phases, particularly in Phase 3. | Means and *SDs* were reported as complementary descriptive indicators; the primary interpretation was based on developmental patterns and non-parametric analyses. |
| Association between tool diversity and pedagogical creativity (*H2*) | Diversity of GenAI tools × pedagogical creativity | Phase 1: $n = 39$; Phase 2: $n = 52$; Phase 3: $n = 45$; overall: $n = 136$ | The relationship was examined as an association between ordinal and discrete variables, without assuming strict linearity or normality. | Spearman's rank-order correlation coefficient (Spearman's ρ) was calculated for each PAR phase and for the overall dataset. |
| Differences by pedagogical GenAI use profile | Pedagogical creativity across four GenAI use profiles | Phase 3: $n = 45$ | The dependent variable was ordinal, and the four GenAI use profiles exhibited unequal group sizes and heterogeneous dispersion. | Kruskal–Wallis H test was used to compare creativity across profiles, with effect size estimated using $\varepsilon^2$ ($H(3) = 18.34$, $p < 0.001$, $\varepsilon^2 = 0.37$). |
| Final satisfaction survey | Overall score and Likert-scale dimensions | $n = 83$ | Aggregated Likert-scale scores; the analysis focused on descriptive statistics and internal consistency rather than inferential testing. | Descriptive statistics, 95% confidence intervals, Cronbach's alpha, and McDonald's omega. |

#### 2.5.2. Units of Analysis and Sample Sizes

Because the study integrated evidence collected across different PAR phases and research instruments, the unit of analysis varied according to the objective of each analysis. This distinction is essential for interpreting the findings, as the initial diagnostic survey was administered to individual participants, whereas the analytical rubric was applied to the final teaching and learning proposals, which were produced either individually or collaboratively depending on the organisation of each learning activity. Similarly, the

Creativity Scale for GenAI Use was applied to the learning artefacts generated throughout the intervention.

Of the 178 artefacts evaluated, 136 contained complete information on both pedagogical creativity and GenAI tool diversity and were included in the correlational analysis. The remaining 42 artefacts lacked a valid score for at least one of these two variables and were therefore excluded only from this specific analysis; they remained available for the descriptive analyses for which complete information was present. No values were imputed. Because the mechanism underlying these incomplete data could not be formally established, the possibility of selection bias cannot be entirely excluded, and the correlational findings should therefore be interpreted with caution.

To facilitate the interpretation of the results, Table 4 summarises the correspondence between each research instrument, its unit of analysis, the associated sample size, and its analytical purpose.

**Table 4.** Units of Analysis and Sample Sizes by Research Instrument.

| Research Instrument/Data Source | Study Phase | Unit of Analysis | n | Analytical Purpose |
|---|---|---|---|---|
| Initial diagnostic survey | Beginning of the intervention | Individual participants | 83 | To establish the cohort's baseline profile, including prior knowledge of UDL, familiarity with GenAI, and perceived learning needs. |
| Group reflective journal (Padlet) | Throughout the intervention | Group journal entries (team-based) | Up to three entries per team | To examine participation, engagement, pedagogical decision-making, and reflective practice throughout the PAR phases. |
| Learning activities and interim artefacts | Throughout the intervention | Team-generated learning artefacts | 178 artefacts | To examine the development of pedagogical creativity in the use of GenAI across the three PAR phases. |
| Creativity Scale for GenAI Use | Throughout the intervention and at the final stage | Learning artefacts containing codable evidence of GenAI use | 178 artefacts; 136 complete cases for correlational analyses | To assess levels of pedagogical creativity and examine their association with the diversity of GenAI tools used. |
| Analytical Rubric for Teaching and Learning Proposals | Final stage | Final teaching and learning proposals | 80 proposals | To evaluate the pedagogical quality, inclusiveness, collaborative design, and feasibility of the final teaching and learning proposals. |
| Final satisfaction survey | Final stage | Individual participants | 83 | To examine participants' evaluation of the MIIA-R model, perceived professional transfer, and overall satisfaction with the intervention. |
| Open-ended responses from the final survey | Final stage | Valid textual responses | Varies by survey question | To identify emergent themes relating to the strengths, challenges, and perceived transfer of the MIIA-R model. |

#### 2.5.3. Independent Second Assessment and Inter-Rater Reliability

To reduce the risk of confirmation bias arising from the dual instructor/assessor role inherent in participatory designs, an independent second rater was incorporated. This rater was external to the teaching team and had no involvement whatsoever in the intervention. The second rater remained blinded to the study hypotheses, to the phase from which each piece of evidence originated, and to the scores initially assigned.

A stratified subsample was selected: 25 of the 80 final proposals (31.3%) and 54 of the 178 learning artefacts (30.3%). Stratification was carried out by PAR phase and by performance level (low, intermediate, and high), ensuring the presence of cases from all empirically observed rubric and creativity-scale levels. Selection within each stratum was random.

In accordance with the established procedure, the second rater received only the anonymised artefacts, the analytical rubric, and the creativity scale with its descriptors, together with a familiarisation session on the instruments that included no information about previous results or about the study hypotheses. Scores were recorded independently, and no discussion of individual cases was permitted before agreement was computed.

For the agreement analysis, the intraclass correlation coefficient (ICC) was calculated for the overall and averaged rubric scores, using the two-way random-effects, absolute-agreement form, ICC(2,1). For the ordinal levels (rubric 1–5 and creativity scale N1–N5), Cohen's weighted kappa with quadratic weights was calculated. Conventional interpretation criteria were adopted: ICC $\geq$ 0.75 indicates good agreement and $\geq$0.90 excellent agreement; $\kappa$ between 0.61 and 0.80 indicates substantial agreement and >0.80 almost perfect agreement.

### *2.6. Data Collection and Analysis Procedures*

Data collection was organised into three sequential stages aligned with the implementation of the MIIA-R. During the initial stage, participants completed the diagnostic survey to identify their prior knowledge of UDL and GenAI, previous experience with these approaches, and perceived learning needs. Throughout the intervention, corresponding to the three PAR phases, evidence was gathered from assessed learning activities, group reflective journals, and learning artefacts to monitor participation, engagement, and the progressive development of participants' teaching and learning proposals. During the final stage, the analytical rubric, the Creativity Scale for GenAI Use, and the final satisfaction survey were administered to evaluate the quality of the proposals, pedagogical creativity, and participants' perceptions of the intervention.

Rather than adopting a conventional pretest–posttest design based on identical measures, the study followed a principle of analytical correspondence between each stage of the intervention and the research instruments employed. The diagnostic survey established the contextual baseline, the reflective journals and learning artefacts documented the implementation process, and the analytical rubric, Creativity Scale for GenAI Use, public challenge presentations, and final satisfaction survey informed the evaluation of the intervention outcomes. This approach ensured coherence between the research objectives, data collection procedures, and the interpretation of the findings. Table 5 summarises the correspondence between the analytical dimensions, stages of data collection, and research instruments.

Quantitative analyses included descriptive statistics (means, standard deviations, frequencies, percentages, and 95% confidence intervals), the Wilcoxon signed-rank test to compare the overall rubric score against the predefined functional threshold (M = 3.0), Spearman's rank-order correlation coefficient ($\rho$), and the Kruskal–Wallis H test to compare pedagogical creativity across GenAI use profiles. The between-profile analysis focused on the overall group comparison, reporting the H statistic, degrees of freedom, $p$ values, and

effect sizes ($\varepsilon^2$). When the omnibus Kruskal–Wallis test indicated statistically significant differences, pairwise post hoc comparisons were conducted using Dunn's test with Bonferroni correction for the six possible comparisons. The analysis reported the H statistic, degrees of freedom, $p$-values, effect sizes ($\varepsilon^2$), and Bonferroni-adjusted $p$-values for the corresponding pairwise contrasts. For the final satisfaction survey, 95% confidence intervals for Cronbach's alpha and McDonald's omega were estimated using bootstrap resampling (1000 iterations). Confidence intervals for Spearman's ρ were calculated using the Fisher z transformation.

**Table 5.** Data Collection and Analysis Procedures by Study Phase.

| PAR Phase | Objective | Evidence Collected | Analysis Performed | Analytical Output |
|---|---|---|---|---|
| Initial phase | To establish the cohort's contextual baseline | Initial diagnostic survey: prior knowledge of UDL and GenAI, and perceived learning needs | Frequencies, percentages, and descriptive analysis | Contextual justification of the MIIA-R model and the sequencing of UDL—GenAI |
| Process phase | To examine participation, engagement, and the development of teaching and learning proposal design | Padlet entries (1–3), assessed learning activities, and interim learning artefacts | Inductive thematic analysis, temporal triangulation, and phase-by-phase monitoring | Reconstruction of the pedagogical process and participants' design decision-making |
| Final phase | To assess proposal quality, pedagogical creativity, and participants' final perceptions | Analytical rubric, Creativity Scale for GenAI Use, and final satisfaction survey | Means and *SDs*; descriptive comparison with the functional threshold; Spearman's rank-order correlation coefficient (Spearman's ρ); Kruskal–Wallis $H$ test; effect size ($\varepsilon^2$) | Contextual evaluation of the MIIA-R; creativity trajectory; hypothesis testing |

To ensure consistency in the control of Type I error, the Bonferroni correction was applied exclusively to the pairwise post hoc comparisons (Dunn's test), where the multiplicity of contrasts (six comparisons) increased the risk of inflating the error rate. The omnibus tests (Kruskal–Wallis, Wilcoxon, Spearman correlations) were interpreted using a single significance level of $\alpha = 0.05$ per hypothesis and therefore did not require additional correction. This differentiation between omnibus tests and multiple comparisons follows standard recommendations in the methodological literature (Field, 2018; Lee & Lee, 2018).

Qualitative data from the group reflective journals and the open-ended survey responses were analysed using inductive thematic analysis following Braun and Clarke (2006). The unit of coding was the smallest text segment conveying a complete meaning. Two researchers independently conducted the initial coding and subsequently refined the coding framework through discussion and consensus, supported by a traceability matrix linking text segments, codes, and themes. Category frequencies were calculated as the proportion of valid responses containing each theme; because individual responses could contribute to multiple themes, frequencies were not mutually exclusive. As the purpose of this phase was to develop consensual, pedagogically informed interpretations rather than to estimate psychometric agreement, no statistical measure of intercoder reliability (e.g., Cohen's κ or the intraclass correlation coefficient [ICC]) was calculated. Instead, the trustworthiness of the qualitative analysis was strengthened through researcher triangulation, source triangulation, and systematic consideration of discrepant evidence.

Beyond the three themes reported above, the inductive thematic analysis of the reflective journals identified two additional interpretive patterns that enrich the reading of the quantitative findings. A fourth theme, the tension between technological efficiency and pedagogical deliberation, revealed that teams perceived GenAI as a productivity catalyst that could, however, shorten collective deliberation on didactic decisions. One group documented this tension explicitly: "*The AI saved us a great deal of time, but we realised we were accepting its first answer without discussing it; we decided to impose a rule on ourselves: every AI proposal had to be discussed and rewritten by the group before being included in the design*" (Group 4, Padlet C2). A fifth theme, the emergence of critical awareness of pedagogical authorship, documented how the peer validation process at the Challenge Fair triggered an explicit revision of the limits of the technology: "*When we defended our proposal in front of our classmates, their questions forced us to justify every decision; that was when we understood that AI generates content, but pedagogical design remained our responsibility*" (Group 11, Padlet C3). These themes converge with the non-linear trajectory observed in the creativity scale and provide interpretive evidence for the pedagogical selectivity effect discussed below: the decrease in creativity scores in the final phase does not reflect a loss of capacity, but a deliberate concentration on viability and on the public defence of the teams' own decisions, to the detriment of further technological exploration.

It should be noted that the "situated pedagogical facilitator" framing is intended in an exclusively associative and contextual sense, not a causal one, given the single-group design of the study.

### 2.7. Researcher Positionality and Reflexivity

As this study was conducted within a Participatory Action Research (PAR) framework in an authentic virtual higher education setting, the researchers did not occupy an external position with respect to the phenomenon under investigation. Instead, they were actively involved in the design, facilitation, and evaluation of the learning experience. Consequently, the researchers explicitly acknowledge their epistemological and pedagogical proximity to the object of study. While this position enabled the generation of situated knowledge, it also introduced potential risks of confirmation bias, interpretative desirability, and overly favourable interpretations of the findings.

Three principal sources of potential bias were identified: (a) the researchers' theoretical commitment to inclusive education, pedagogical innovation, and the educational use of GenAI; (b) the potential influence of their facilitative teaching role on participants' engagement and responses; and (c) the possibility of interpreting the findings as unequivocal support for the MIIA-R despite differential levels of performance across its dimensions. To mitigate these risks, the study incorporated methodological triangulation across data sources and research instruments, independent initial coding followed by analytical consensus between researchers, a traceability matrix linking evidence to codes and themes, and systematic consideration of discrepant or non-confirmatory evidence, particularly in dimensions whose performance fell below the predefined functional threshold.

Within this study, reflexivity was understood not as an autobiographical exercise but as a criterion of methodological rigour intended to make explicit the conditions under which knowledge was produced and to support transparent, cautious, and auditable interpretations. Consistent with the principles of PAR, the findings are interpreted as contextually situated and robust within the educational setting examined rather than as neutral or universally generalisable without replication and evaluation in other educational contexts (Berger, 2015; Herr & Anderson, 2015; Creswell & Creswell, 2018).

### *2.8. Declaration on the Use of Generative Artificial Intelligence*

The GenAI tools examined in this study were used exclusively by participants as part of the educational intervention. The authors did not use GenAI to generate research data, modify the dataset, perform statistical analyses, or produce the empirical findings.

During the preparation of this manuscript, ChatGPT (OpenAI, web version, accessed 20 July 2026) was used solely to support language revision, including improvements to wording, clarity, grammar, and stylistic consistency. Manus AI (web version, accessed 20 July 2026) was used to assist in the preparation of figures based exclusively on data, analytical results, and specifications previously produced and verified by the authors. No AI-generated data, analyses, or empirical findings were incorporated into the study.

All AI-assisted outputs were critically reviewed, edited, and validated by the authors, who assume full responsibility for the content of the manuscript. All methodological, analytical, and interpretative decisions remained exclusively under the authors' responsibility.

## 3. Results

The results are organised into five analytical sections corresponding to the study objectives: (1) the cohort's baseline characteristics; (2) participation and engagement throughout the intervention; (3) the quality of the inclusive teaching and learning proposals; (4) pedagogical creativity in the use of GenAI; and (5) the relationship between GenAI tool diversity and pedagogical creativity. Together, these analyses provide an integrated account of the principal dimensions of the research problem. Figure 3 summarises the analytical framework underpinning the presentation of the findings.

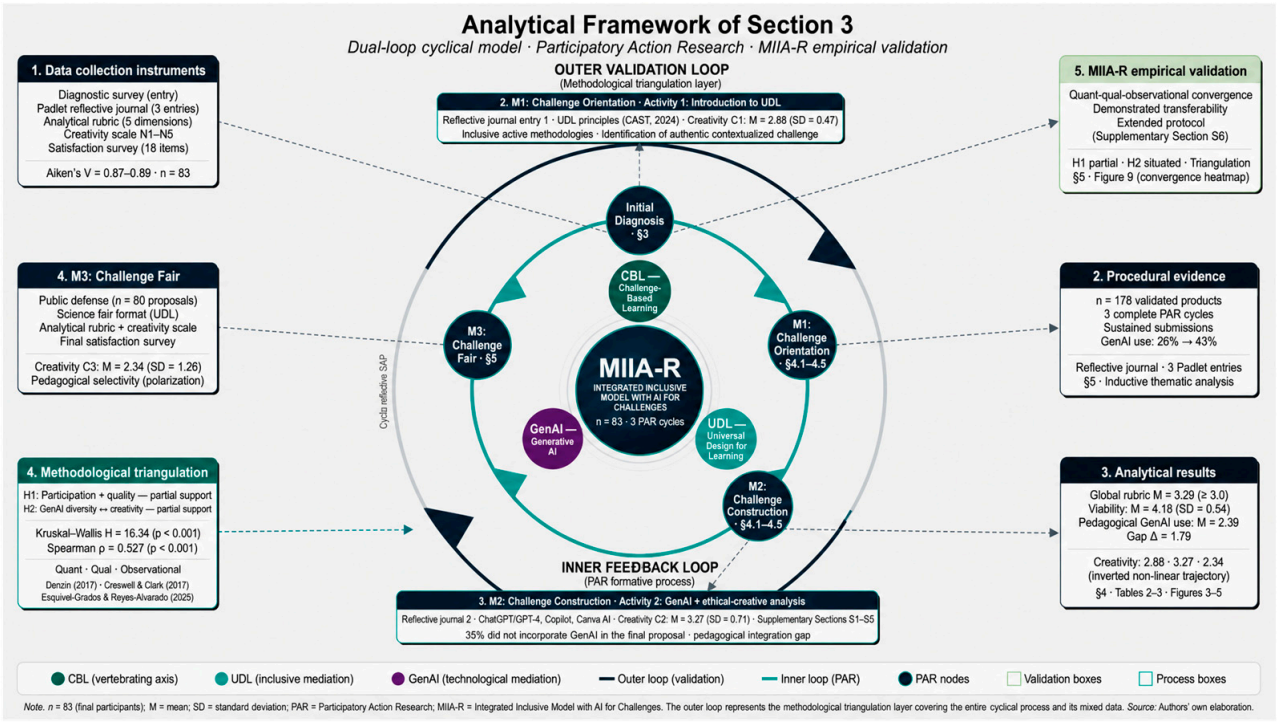



**Figure 3.** Analytical Framework of Section 3. References cited in the figure: CAST (2024), Denzin (2017), Creswell and Clark (2017), and Esquivel-Grados and Reyes-Alvarado (2025).

Figure 3 illustrates the dual-loop structure guiding the empirical evaluation of the MIIA-R. The outer loop represents the methodological triangulation integrating quantitative, qualitative, and observational evidence, whereas the inner loop depicts the formative progression across the three PAR phases (Challenge Orientation, Challenge Development,

and Challenge Showcase). At the centre of the model, CBL provides the pedagogical structure, UDL supports inclusive learning processes, and GenAI functions as the technological mediation underpinning the intervention.

### *3.1. Baseline Characteristics of the Cohort: Initial Diagnostic Assessment*

The initial diagnostic survey (n = 83) revealed a heterogeneous baseline, providing the contextual starting point for the intervention. Only 37.3% of participants reported being familiar with UDL, whereas 92.8% recognised the pedagogical potential of GenAI, and 86.7% expressed a need for specific training integrating UDL and GenAI. This baseline profile indicates a marked asymmetry between participants' strong interest in GenAI and their limited prior knowledge of the inclusive pedagogical framework required to guide its educational use. Figure 4 summarises these findings.

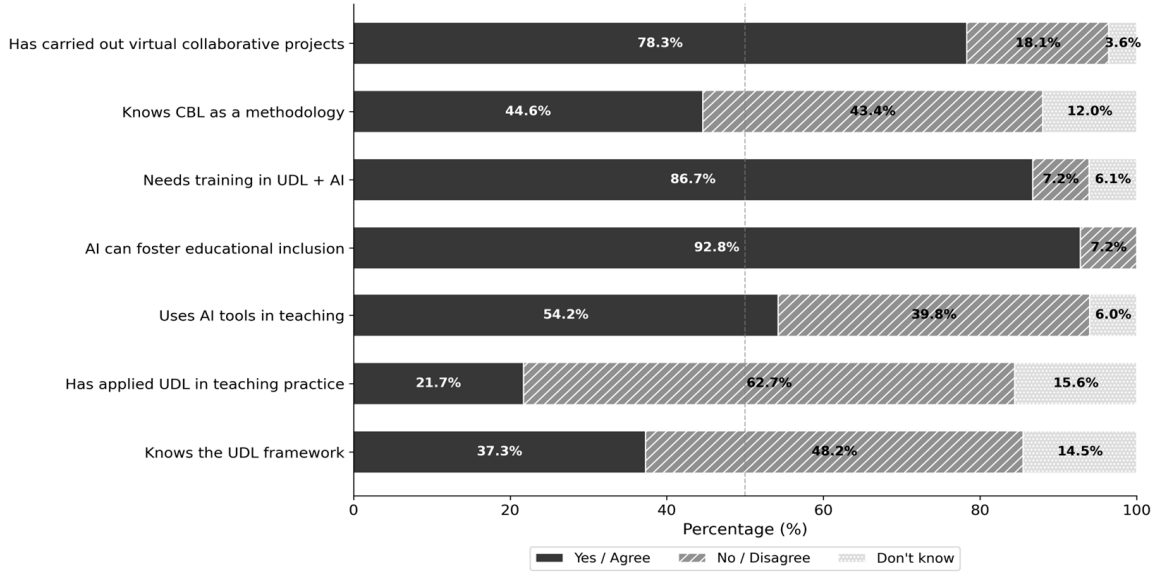



**Figure 4.** Initial diagnosis: prior knowledge and perceptions regarding UDL and GenAI (n = 83). Stacked horizontal bars (Yes/No/Don't know).

The diagnostic survey was designed to characterise the educational context at the outset of the intervention rather than to provide a pretest for subsequent statistical comparison. Accordingly, these findings are interpreted as contextual baseline evidence. The final evaluation subsequently showed a difference between proposal feasibility (M = 4.18) and the pedagogical use of GenAI (M = 2.39; Δ = 1.79), which is reported in the corresponding results section and discussed in Section 4.1.

### *3.2. Participation and Engagement During Challenge Development*

During the second PAR phase, participants developed their teaching and learning proposals using GenAI, building on Activity 1 (UDL-based learning sequences) and extending their work through Activity 2, which combined GenAI-supported instructional design with collaborative ethical and creative analysis. Participation and engagement were documented through the completion of learning activities, evidence recorded in the group reflective journals, and the learning artefacts produced throughout the intervention.

Analysis of the final teaching and learning proposals (n = 80) showed that 35.0% did not incorporate GenAI into their proposals. Among those that did, ChatGPT/GPT-4 was the most frequently used tool (21.2%), followed by a variety of other GenAI tools (37.5%) and Canva AI (2.5%). Figure 5 summarises the distribution of GenAI tools incorporated into the final proposals. Percentages refer to the primary GenAI tool identified in each

proposal; the remaining proposals (~3.8%) reported the combined use of multiple tools and are described in Section Evolution of GenAI Tool Use Profiles.

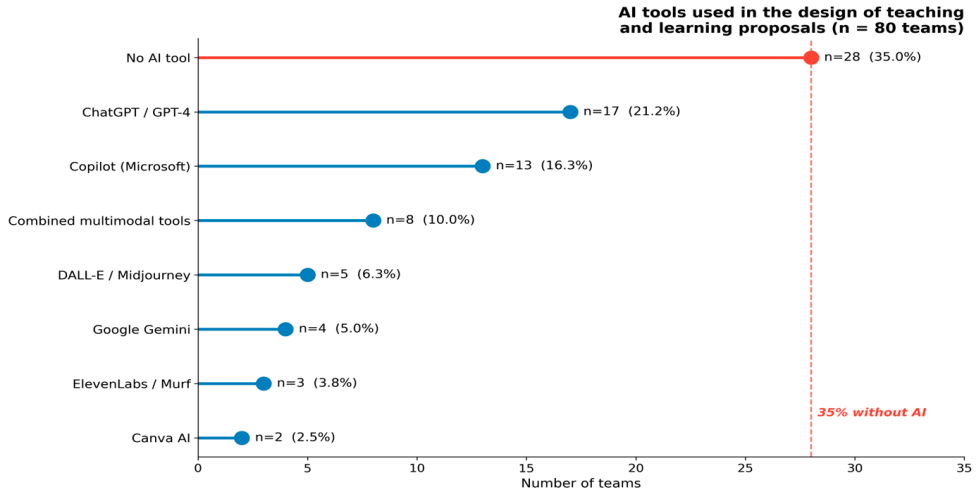



**Figure 5.** GenAI Tools Used in the Teaching and Learning Proposals ($n$ = 80). Lollipop chart displaying the frequency of GenAI tool use across the analysed proposals.

These findings describe heterogeneous patterns of GenAI integration across participants and provide the empirical context for the subsequent analyses of pedagogical creativity and proposal quality.

### *3.3. Quality of the Inclusive Teaching and Learning Proposals at the Challenge Showcase*

The Challenge Showcase constituted the final dissemination and public presentation stage of the intervention. During this phase, the 80 teaching and learning proposals developed in PAR Phase 3 were presented to the academic community, discussed through peer feedback, and evaluated using the analytical rubric. Consistent with the instructional design of the module, the Showcase was organised according to UDL principles, allowing participants to experience inclusive learning both as presenters and as members of the audience.

The analytical rubric ($n$ = 80 proposals) yielded an overall mean score of $M$ = 3.29 ($SD$ = 0.63; 95% CI [3.15, 3.43]), exceeding the predefined functional threshold ($M$ = 3.0). The median score was $Mdn$ = 3.30. A one-sample Wilcoxon signed-rank test indicated that the overall quality of the teaching and learning proposals was significantly higher than the predefined functional threshold (W = 2450, $p < 0.001$). Table 6 and Figure 6 summarise the overall results together with the mean scores obtained for each rubric dimension. Detailed descriptors for each performance level are provided in Supplementary Materials (Section S3).

Across the rubric dimensions, proposal feasibility achieved the highest score (Mdn = 4.20; M = 4.18), whereas the pedagogical use of GenAI obtained the lowest (Mdn = 2.20; M = 2.39). UDL integration approached the predefined functional threshold, while the remaining dimensions exceeded it.

Although the overall rubric score exceeded the predefined functional threshold, this result was primarily driven by proposal feasibility and curricular coherence. The core dimensions related to UDL and the pedagogical integration of GenAI remained below the functional threshold, indicating that the MIIA-R supported the production of feasible instructional proposals but did not consistently achieve deep inclusive and AI-mediated pedagogical integration.

The qualitative evidence derived from the group reflective journals complements these quantitative findings by documenting how participants experienced the design process,

the pedagogical integration of GenAI, and the collaborative learning dynamics during the Challenge Showcase. These findings are presented in the following section as part of the study's methodological triangulation.

**Table 6.** Descriptive Results of the Analytical Rubric for the Final Teaching and Learning Proposals ($n$ = 80).

| Dimension | Weight (%) | M | *SD* | 95% CI | Interpretation |
|---|---|---|---|---|---|
| D1—Curricular Coherence | 20 | 3.62 | 0.61 | [3.48–3.76] | Above the Functional Threshold |
| D2—UDL Integration | 25 | 2.88 | 0.74 | [2.72–3.04] | Close to the Functional Threshold |
| D3—Pedagogical Use of GenAI | 25 | 2.39 | 0.81 | [2.21–2.57] | Below the Functional Threshold |
| D4—Collaborative Work | 15 | 3.58 | 0.68 | [3.43–3.73] | Above the Functional Threshold |
| D5—Proposal Feasibility | 15 | 4.18 | 0.54 | [4.06–4.30] | Above the Functional Threshold |
| Overall (Weighted Score) | 100 | 3.29 | 0.63 | [3.15–3.43] | Above the Functional Threshold |

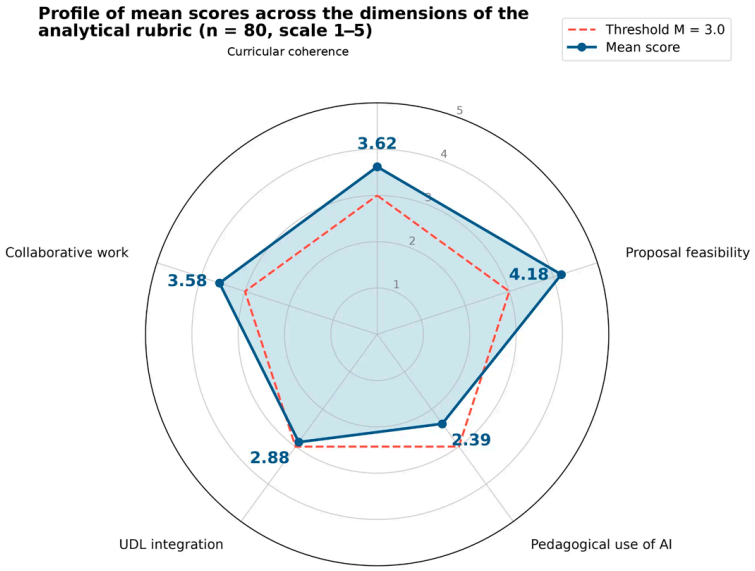



**Figure 6.** Profile of Mean Scores Across the Dimensions of the Analytical Rubric.

#### 3.3.1. Participation and Engagement Observed During the Challenge Showcase: Qualitative Evidence

The Challenge Showcase was conceived as a collaborative peer-learning event inspired by the format of a science fair while incorporating UDL principles throughout its organisation, including physical and digital accessibility, multiple presentation formats, and flexible participation arrangements. The qualitative findings presented below complement the quantitative results by documenting how participants experienced the design, presentation, and collaborative evaluation of their teaching and learning proposals during the final stage of the intervention.

Inductive thematic analysis of the group reflective journals identified three recurring themes concerning the integration of inclusive pedagogy and GenAI: (a) awareness of algorithmic bias, whereby participants recognised that GenAI does not inherently generate accessible or inclusive learning designs; (b) reframing teacher agency, characterised by the deliberate prioritisation of pedagogical judgement over automated content generation; and (c) the value of peer assessment, highlighting the Challenge Showcase as a space for collaborative reflection and professional learning. These themes were identified across multiple groups and reflective entries. The excerpts presented below are representative examples selected to illustrate each theme.

- "*At first, we thought that using ChatGPT would automatically produce a high-quality design, but after revisiting the UDL principles, we realised that AI alone does not guarantee inclusion. We had to rethink the entire proposal.*" (Group 3, Padlet C2)
- "*The Showcase was the moment when we truly understood what designing for everyone means. Seeing the other groups' proposals made us reflect on what was missing from our own.*" (Group 7, Padlet C3)
- "*We used Copilot to generate supporting resources and were surprised that AI did not take student diversity into account. We had to incorporate that perspective ourselves.*" (Group 12, Padlet C2)
- "*The greatest lesson we learned from this module was that technology alone does not create inclusion; it is teachers who make inclusion possible through sound pedagogical judgement.*" (Group 5, Padlet C3)
- "*For the final challenge, we decided not to use generative AI to create the learning content itself but only to review accessibility. It required more work, but the final product genuinely reflected our own pedagogical decisions.*" (Group 9, Padlet C3)

*Note.* All quotations were anonymised in accordance with the study's ethical protocol. Their original wording has been retained to preserve the authenticity of participants' voices.

#### 3.3.2. Inter-Rater Reliability of the Analytical Rubric and the Creativity Scale

The independent re-evaluation of the stratified subsample showed good and statistically significant agreement. For the overall rubric score, the intraclass correlation coefficient was ICC(2,1) = 0.84 (95% CI [0.71, 0.92], $p < 0.001$). For the Creativity Scale (N1–N5), the intraclass estimate also showed good agreement (ICC = 0.82, 95% CI [0.70, 0.90]). Across dimensions, values ranged from ICC = 0.76 for the pedagogical use of GenAI to ICC = 0.89 for proposal feasibility, indicating that the dimensions requiring greater interpretative judgement (D2 and D3) showed somewhat lower agreement, although still within acceptable limits. For the ordinal levels, Cohen's weighted kappa was $\kappa w = 0.74$ (95% CI [0.62, 0.85]) for the rubric and $\kappa w = 0.78$ (95% CI [0.66, 0.87]) for the creativity scale (N1–N5), values consistent with substantial agreement.

These results do not entirely eliminate the risk of bias associated with the dual instructor/assessor role, but they support the conclusion that the scores were not decisively dependent on the individual rater and strengthen the credibility of the estimates presented. Discrepancies were concentrated precisely in dimensions D2 and D3, which is consistent with the greater interpretative demands of UDL integration and of the pedagogical use of GenAI (Table 7).

**Table 7.** Inter-rater agreement on the stratified subsample re-evaluated by an independent blinded rater.

| Dimension | Unit Re-Evaluated | n Re-Evaluated | ICC(2,1) | 95% CI | Weighted κ |
|---|---|---|---|---|---|
| Analytical rubric—overall score | Final proposals | 25 (31.3%) | 0.84 | [0.71, 0.92] | 0.74 |
| D1—Curricular coherence | Final proposals | 25 | 0.87 | [0.74, 0.94] | 0.79 |
| D2—UDL integration | Final proposals | 25 | 0.80 | [0.63, 0.90] | 0.71 |
| D3—Pedagogical use of GenAI | Final proposals | 25 | 0.76 | [0.57, 0.88] | 0.68 |
| D4—Collaborative work | Final proposals | 25 | 0.85 | [0.71, 0.93] | 0.77 |
| D5—Proposal feasibility | Final proposals | 25 | 0.89 | [0.78, 0.95] | 0.82 |
| Creativity Scale for GenAI Use (N1–N5) | Learning artefacts | 54 (30.3%) | 0.82 | [0.70, 0.90] | 0.78 |

Note. ICC(2,1) = two-way random effects intraclass correlation coefficient with absolute agreement; weighted κ = Cohen's kappa with quadratic weights. The independent rater was external to the teaching team and blinded to the study hypotheses and to the original scores. Sampling was stratified by PAR phase and performance level.

### *3.4. Pedagogical Creativity in the Use of GenAI*

The Creativity Scale for GenAI Use was applied to a total of 178 learning artefacts to examine patterns of pedagogical creativity across the three PAR phases. The observed distributions showed relatively homogeneous creativity levels during Phase 1, an increase in creativity scores during Phase 2, and a bimodal distribution accompanied by a lower central tendency during Phase 3. Figure 7 summarises these patterns using measures of central tendency and dispersion.

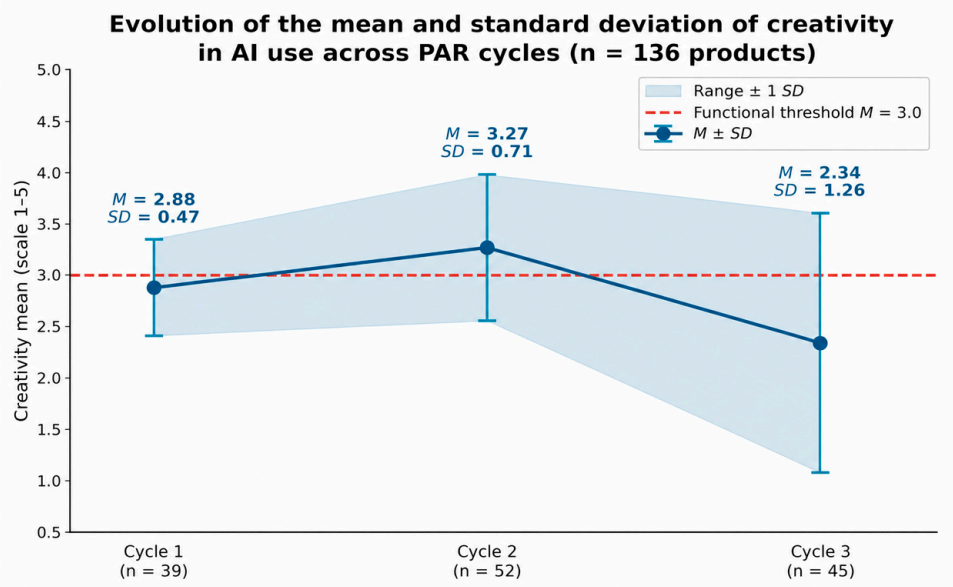



**Figure 7.** Mean (*M*) and standard deviation (*SD*) of pedagogical creativity in GenAI use across the three PAR phases. The observed temporal fluctuation (2.88 → 3.27 → 2.34) and the increase in *SD* (0.47 → 0.71 → 1.26) reflect greater dispersion in the final cycle.

Because the Creativity Scale is an ordinal five-level instrument (N1–N5), mean values and standard deviations are reported only as complementary descriptive indicators. The interpretation of the findings is based primarily on the distribution of creativity scores across the three PAR phases, the observed variability in the data, and the non-parametric analyses presented below, rather than on assumptions of interval-level measurement.

The descriptive results indicate a progressive diversification of pedagogical creativity throughout the intervention. Whereas creativity levels were relatively homogeneous during Phase 1, greater variability emerged in Phase 2 as participants began to integrate GenAI into the design of their teaching and learning proposals. During Phase 3, creativity scores became more heterogeneous, reflecting differentiated patterns in the pedagogical use of GenAI during the final development of the proposals. The inferential analyses presented in the following sections examine these patterns in greater detail.

#### Evolution of GenAI Tool Use Profiles

Figure 8 illustrates the evolution of participants' GenAI tool use profiles across the three PAR phases. The proportion of proposals relying exclusively on a single GenAI tool decreased from 44% in Phase 1 to 33% in Phase 3, whereas the use of multimodal combinations increased from 26% to 43% over the same period. Concurrently, participants incorporated a broader range of GenAI functionalities, including accessibility support, content revision, information organisation, and pedagogical refinement, in addition to content generation.

These findings describe a progressive diversification of GenAI use throughout the intervention and provide the empirical context for the subsequent analyses of pedagogical creativity and the relationship between GenAI tool diversity and creativity.

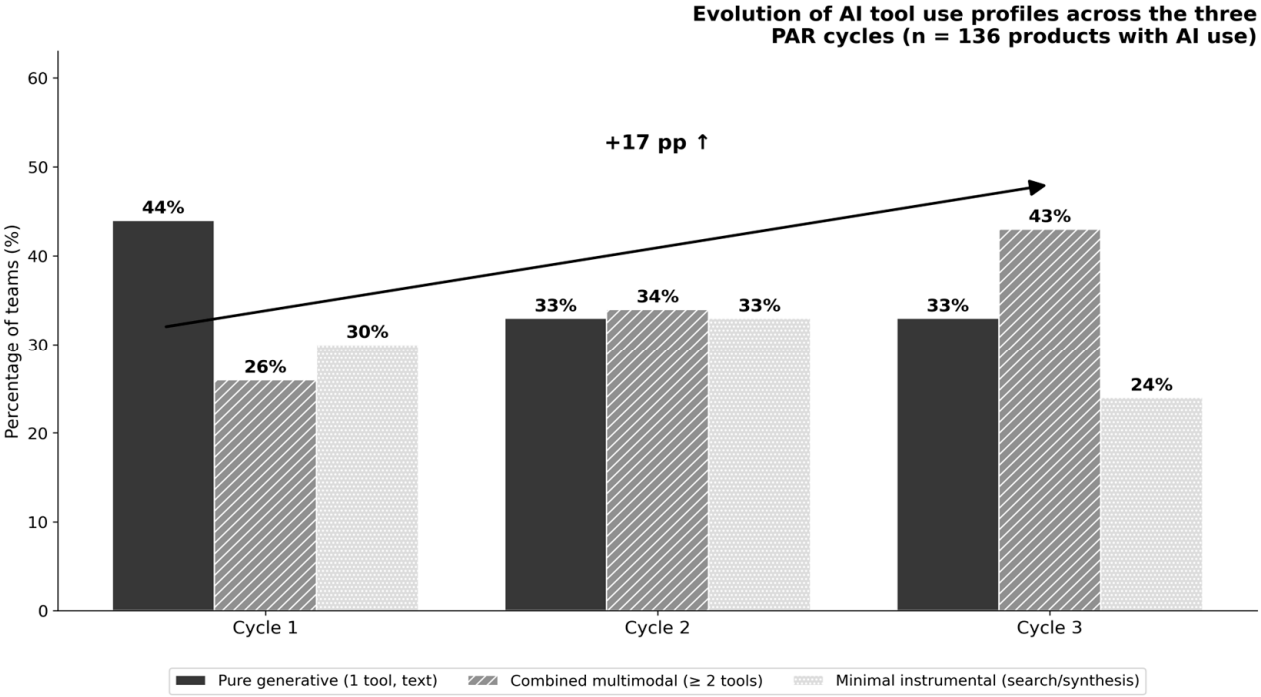



**Figure 8.** Evolution of GenAI Tool Use Profiles Across the Three PAR Phases. Grouped bar chart of GenAI tool use profiles across the three PAR phases. The arrow indicates a 17-percentage-point increase in combined multimodal GenAI tool use, from 26% in Cycle 1 to 43% in Cycle 3.

### *3.5. Association Between GenAI Tool Diversity and Pedagogical Creativity*

Spearman's rank-order correlation coefficient (ρ) was used to examine the association between GenAI tool diversity and pedagogical creativity, both across the individual PAR phases and for the complete dataset. GenAI tool diversity was operationalised as the number of distinct GenAI tools or functions employed during the design process. Table 8 summarises the correlation coefficients obtained for each PAR phase together with the overall pattern of association.

**Table 8.** Spearman's Rank-Order Correlation Between GenAI Tool Diversity and Pedagogical Creativity Across the Three PAR Phases.

| PAR Phase | n | Spearman's ρ | *p* | Interpretation |
|---|---|---|---|---|
| C1 | 39 | 0.288 | 0.075 | Weak positive association, not statistically significant. |
| C2 | 52 | 0.211 | 0.132 | Weak positive association, not statistically significant. |
| C3—Final Challenge | 45 | 0.527 | <0.001 | Moderate positive association, statistically significant |
| Global | 136 | 0.247 | <0.01 | Weak positive association, statistically significant. |

Note. ρ = Spearman's rank-order correlation coefficient. GenAI tool diversity was operationalised as the number of different GenAI tools or distinct GenAI functions documented in each learning artefact. Pedagogical creativity was assessed using the five-level ordinal Creativity Scale (Supplementary Materials, Section S1).

To complement the correlational analyses, pedagogical creativity was also compared across four pedagogical GenAI use profiles. Rather than classifying participants according to the specific tools employed, the analysis focused on patterns of pedagogical technology appropriation reflected in the learning artefacts. The four profiles comprised: (a) no or

minimal GenAI use; (b) generative-only use, primarily involving text or resource generation; (c) guided instrumental use, in which GenAI supported specific pedagogical tasks such as instructional design, revision, or accessibility enhancement; and (d) multimodal integrated use, characterised by the purposeful combination of multiple GenAI tools or functions to generate, review, adapt, and refine teaching and learning proposals. This classification enabled comparisons of pedagogical creativity across different patterns of GenAI use within the same participant cohort without establishing experimental or control groups.

GenAI tool diversity was coded independently of pedagogical creativity. Whereas tool diversity was operationalised exclusively as the number of distinct GenAI tools or functions employed by each team, the Creativity Scale assessed the originality, inclusive pedagogical relevance, and instructional depth of the final teaching and learning proposals, irrespective of the number of tools used. This methodological separation minimised the risk of analytical circularity in the correlational analyses.

To facilitate interpretation, 95% confidence intervals were calculated for the principal Spearman correlation coefficients. During PAR Phase 3 (Final Challenge), a moderate positive association was observed between GenAI tool diversity and pedagogical creativity ($\rho = 0.527$, 95% CI [0.287, 0.701], $p < 0.001$). Across the complete dataset, the association was weaker but remained statistically significant ($\rho = 0.247$, 95% CI [0.067, 0.414], $p < 0.01$).

Differences in Pedagogical Creativity Across GenAI Use Profiles: Kruskal–Wallis Analysis

To complement the correlational analyses, a Kruskal–Wallis H test was conducted to compare levels of pedagogical creativity across the four pedagogical GenAI use profiles identified within the same participant cohort: (a) no or minimal GenAI use, (b) generative-only use, (c) guided instrumental use, and (d) multimodal integrated use.

As shown in Table 9, statistically significant differences were observed across the four profiles during PAR Phase 3, $H(3) = 18.34$, $p < 0.001$, $\varepsilon^2 = 0.37$. The corresponding effect size was large, indicating substantial differences in pedagogical creativity among the four GenAI use profiles. Descriptively, the multimodal integrated use profile achieved the highest mean creativity score (M = 3.6), whereas the generative-only use profile obtained the lowest (M = 1.8).

**Table 9.** Kruskal–Wallis H Test Comparing Pedagogical Creativity Across Pedagogical GenAI Use Profiles in PAR Phase 3.

| Comparison | *n* | GenAI Use Profiles Compared | Test Statistic | df | *p* | Effect Size ($\varepsilon^2$) | Interpretation |
|---|---|---|---|---|---|---|---|
| Overall comparison across GenAI use profiles | 45 | Four pedagogical GenAI use profiles | $H = 18.34$ | 3 | <0.001 | $\varepsilon^2 = 0.37$ | Statistically significant differences in pedagogical creativity with a large effect size. |

Note. H = Kruskal–Wallis test statistic; $\varepsilon^2$ = effect size.

Pairwise comparisons were conducted using Dunn's test with Bonferroni correction for six comparisons. Bonferroni-adjusted *p*-values were evaluated at $\alpha = 0.05$, equivalent to comparing the unadjusted *p*-values with $\alpha/6 = 0.0083$. In addition, and in accordance with the ordinal nature of the variable, the median and interquartile range are reported for each profile, with means and standard deviations retained as complementary information (Field, 2018; Lee & Lee, 2018).

The pairwise comparisons indicated that the overall differences were concentrated in the multimodal integrated use profile. This profile obtained significantly higher pedagogical creativity scores than the generative-only use profile ($z = -3.87$, adjusted $p = 0.001$) and

than the no or minimal use profile (z = −3.12, adjusted $p = 0.011$). By contrast, no statistically significant differences were observed between no or minimal use and generative-only use (z = 0.42, adjusted $p = 1.000$), between guided instrumental use and multimodal integrated use (z = −2.05, adjusted $p = 0.242$), or between generative-only use and guided instrumental use (z = −2.61, adjusted $p = 0.054$), although this last comparison approached the threshold of significance.

Accordingly, the overall effect detected should not be interpreted as a linear progression across the four profiles, but rather as a principal contrast between the multimodal and intentional integration of GenAI and its more elementary uses. This interpretation is consistent with understanding GenAI as a situated pedagogical facilitator rather than as an autonomous explanatory factor (Tables 10 and 11).

**Table 10.** Descriptive statistics by pedagogical GenAI use profile.

| Profile | *n* | *M* | *SD* | *Mdn* | *IQR* |
|---|---|---|---|---|---|
| P1—No or minimal GenAI use | 8 | 2.00 | 0.53 | 2.00 | 1.75–2.25 |
| P2—Generative-only use | 11 | 1.80 | 0.60 | 2.00 | 1.50–2.00 |
| P3—Guided instrumental use | 14 | 2.86 | 0.66 | 3.00 | 2.25–3.00 |
| P4—Multimodal integrated use | 12 | 3.60 | 0.79 | 3.50 | 3.00–4.00 |

Note. M = mean; SD = standard deviation; Mdn = median; IQR = interquartile range. Medians and IQRs are reported as the primary descriptive statistics given the ordinal nature of the creativity scale; means and SDs are reported as complementary information.

**Table 11.** Dunn pairwise post hoc comparisons with Bonferroni correction (PAR Phase 3, n = 45).

| Pairwise Comparison (Dunn) | *z* | *p* (Unadjusted) | *p* (Bonferroni-Adjusted) | Decision |
|---|---|---|---|---|
| P1 vs. P2 | 0.42 | 0.675 | 1.000 | Not significant |
| P1 vs. P3 | −1.95 | 0.051 | 0.307 | Not significant |
| P1 vs. P4 | −3.12 | 0.002 | 0.011 | Significant |
| P2 vs. P3 | −2.61 | 0.009 | 0.054 | Not significant |
| P2 vs. P4 | −3.87 | <0.001 | 0.001 | Significant |
| P3 vs. P4 | −2.05 | 0.040 | 0.242 | Not significant |

Note. Pairwise comparisons were conducted using Dunn's test with Bonferroni correction for six comparisons. Bonferroni-adjusted *p*-values were evaluated at $\alpha = 0.05$, equivalent to comparing the unadjusted *p*-values with $\alpha/6 = 0.0083$.

No statistically significant association between GenAI tool diversity and pedagogical creativity was observed during the initial PAR phases (Phase 1: $\rho = 0.288$, $p = 0.075$; Phase 2: $\rho = 0.211$, $p = 0.132$). By contrast, during PAR Phase 3 (Final Challenge), a moderate positive association was observed ($\rho = 0.527$, $p < 0.001$).

Given the non-experimental, single-group design of the study, these findings describe differences associated with distinct pedagogical GenAI use profiles and should not be interpreted as evidence of causal relationships.

### 3.6. Methodological Triangulation

Drawing on the principles of methodological triangulation (Denzin, 2017; Creswell & Clark, 2017; Esquivel-Grados & Reyes-Alvarado, 2025), quantitative and qualitative evidence was integrated into a five-dimensional analytical matrix linking the contextual baseline, participants' engagement throughout the intervention, the quality of the teaching and learning proposals, and pedagogical creativity in the use of GenAI.

Consistent with the concurrent mixed-methods PAR design described in the Methods section, triangulation was conceived not merely as a means of corroborating findings but as an integrative procedure for generating meta-inferences regarding the contextual evaluation

of the MIIA-R. The combined analysis of the analytical rubric, the Creativity Scale, the group reflective journals, the open-ended survey responses, and the final satisfaction survey indicated that participants' engagement was associated not with the use of GenAI per se but with the way in which it was integrated with UDL principles and authentic challenge-based learning activities. Within this framework, GenAI tool diversity acquired pedagogical relevance when linked to instructional design decisions, critical review, inclusive adaptation, and the refinement of teaching and learning proposals.

Overall, the triangulated evidence provides convergent, partial support for Hypothesis 1 and context-dependent correlational support for Hypothesis 2. The resulting meta-inference indicates that GenAI functioned as a situated pedagogical facilitator rather than as an autonomous explanatory factor, with its contribution to pedagogical creativity and proposal quality depending on pedagogical mediation, inclusive instructional scaffolding, and participants' engagement throughout the learning process. Table 12 summarises the integration of evidence supporting these meta-inferences.

**Table 12.** Joint Display Integrating Evidence for Hypotheses H1 and H2.

| Hypothesis/ Component | Quantitative Evidence | Qualitative Evidence | Integrated Meta-Inference | Interpretative Judgement |
|---|---|---|---|---|
| *H1a.* Sustained participation | n = 83 participants; 178 learning artefacts produced across the PAR phases; sustained participation throughout the intervention | Reflective journals documented continuous engagement, pedagogical decision-making, and progressive refinement of the proposals | Participation was sustained at a level sufficient to support the learning pathway; however, participation intensity alone did not necessarily translate into deeper pedagogical engagement | Convergent support |
| *H1b.* Engagement and reflective practice | Sustained learning activity across PAR Phases 1–3; positive evaluation of the learning experience; high satisfaction and positive Net Promoter Score (NPS) | Reflective journals (Padlet) and open-ended survey responses revealed critical reflection on the limitations of GenAI, the need to adapt learning resources, and the perceived professional value of the intervention | Participant engagement was consistently high and accompanied by pedagogical reflection, although the depth of engagement varied across teams | Moderate convergent support |
| *H1c.* Functional quality of the teaching and learning proposals | Overall analytical rubric score: M = 3.29, exceeding the predefined functional threshold (M = 3.0) | Qualitative evidence indicated that the proposals were perceived as feasible and applicable within authentic educational settings | The MIIA-R model supported the development of pedagogically functional and contextually applicable teaching and learning proposals within the educational setting examined | Convergent support based on the feasibility of the proposals and their curricular coherence |
| *H1d.* Internal constraints on proposal quality | UDL integration: M = 2.88; pedagogical use of GenAI: M = 2.39; proposal feasibility: M = 4.18 | Participants consistently recognised that GenAI alone does not ensure inclusive learning design and requires deliberate pedagogical mediation | Support for Hypothesis 1 is partial, the model supported feasible and curricularly coherent proposals, but it did not consistently achieve functional performance in its two core dimensions: UDL integration and the pedagogical use of GenAI | Partial and limited support: the two core dimensions of the model (UDL integration and pedagogical use of GenAI) did not reach the functional threshold |

**Table 12.** *Cont.*

| Hypothesis/ Component | Quantitative Evidence | Qualitative Evidence | Integrated Meta-Inference | Interpretative Judgement |
|---|---|---|---|---|
| *H2a.* Association between GenAI tool diversity and pedagogical creativity | PAR Phase 3: ρ = 0.527, $p < 0.001$; weaker overall association across the complete dataset | Qualitative evidence indicated that more sophisticated pedagogical uses combined content generation, critical review, and instructional adaptation | Greater GenAI tool diversity was associated with higher levels of pedagogical creativity during the final phase of the intervention; however, the study design does not support causal inference | Context-specific convergent support |
| *H2b.* Differences across pedagogical GenAI use profiles | Kruskal–Wallis: H(3) = 18.34, $p < 0.001$, $\varepsilon^2 = 0.37$. Dunn post hoc comparisons showed that only the multimodal integrated use profile differed significantly from the generative-only and minimal use profiles | Qualitative accounts distinguished between instrumental uses of GenAI and more sophisticated forms of pedagogical appropriation | Differences across GenAI use profiles reinforce the pedagogical relevance of how GenAI is used rather than its mere presence within the instructional design process | Moderate convergent support |
| *H2c.* Temporal trajectory of pedagogical creativity | PAR Phase 1: M = 2.88; PAR Phase 2: M = 3.27; PAR Phase 3: M = 2.34; increased variability observed in PAR Phase 3 | Participants' accounts reflected diverse strategies, including intensive GenAI use, selective use, and the deliberate reduction in GenAI use during the Final Challenge | Pedagogical creativity followed a non-linear trajectory. The decline observed in PAR Phase 3 should be interpreted with caution and may reflect multiple plausible explanations | Partial convergence with an open interpretative pattern |

Note. Hypothesis 1 received partial and limited support: the overall rubric score exceeded the functional threshold, but UDL integration and the pedagogical use of GenAI remained below it. Hypothesis 2 received context-dependent correlational support. The findings reflect associations rather than causal relationships.

This partial support requires an explicit qualification. The overall rubric score exceeded the predefined functional threshold, yet this result was primarily driven by proposal feasibility and curricular coherence, whereas the two core dimensions of the model, UDL integration and the pedagogical use of GenAI, remained below the threshold. Accordingly, the triangulated evidence indicates that the MIIA-R supported the production of feasible and curricularly coherent instructional proposals but did not consistently achieve deep inclusive and AI-mediated pedagogical integration. The overall score should therefore not be read as evidence that all components of the model performed satisfactorily.

The inter-rater agreement estimates reported in Section 3.3.2 (ICC = 0.84; weighted κ = 0.74–0.78) and the Dunn pairwise comparisons reported in the section "Differences in Pedagogical Creativity Across GenAI Use Profiles: Kruskal–Wallis Analysis" strengthen the credibility of these meta-inferences by showing that the rubric scores were not decisively rater-dependent and by identifying which GenAI use profiles actually differed from one another.

Figure 9 complements the triangulation matrix by visually summarising the integration of quantitative and qualitative evidence across the study's five analytical dimensions.

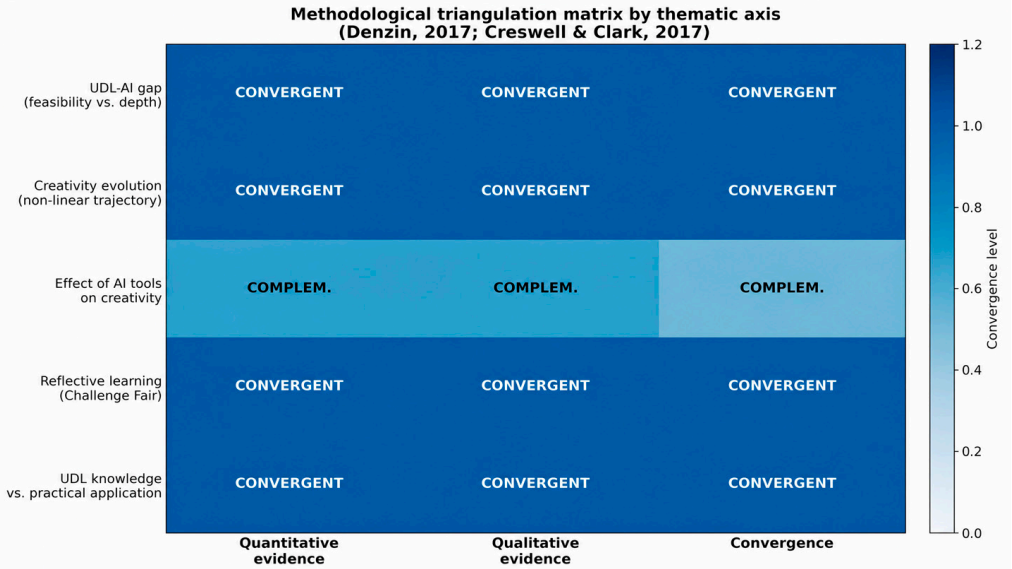



**Figure 9.** Heatmap of methodological triangulation across five analytical dimensions and three sources of evidence. References cited in the figure: Denzin (2017) and Creswell and Clark (2017). Darker colour intensity indicates a higher degree of convergence among the sources of evidence.

## 4. Discussion

The findings are discussed in relation to the theoretical framework underpinning the MIIA-R and current research on the pedagogical integration of GenAI, UDL, and challenge-based learning in teacher education. Consistent with the concurrent mixed-methods PAR design, the discussion is organised around four analytical dimensions derived from the integration of the quantitative and qualitative findings. Particular attention is given to the convergences and tensions identified through methodological triangulation in order to clarify the pedagogical conditions that shaped the meaningful integration of GenAI throughout the intervention.

From this perspective, the findings should be interpreted with appropriate caution. Overall, they suggest the pedagogical value of the MIIA-R while simultaneously revealing important tensions concerning the deeper integration of UDL principles, the pedagogical use of GenAI, and the non-linear relationship between GenAI tool diversity and pedagogical creativity.

### *4.1. The Structural Gap Between Functional Viability and Pedagogical Depth*

This discrepancy should be interpreted as a limitation of the model as implemented rather than as an internal tension within an otherwise successful outcome. Students produced feasible and curricularly coherent instructional proposals, but the MIIA-R did not consistently achieve deep pedagogical integration of UDL (M = 2.88) or of GenAI (M = 2.39). In other words, the model sustained functional instructional production, whereas its distinctive inclusive and AI-mediated core remained below the predefined functional threshold.

This pattern also carries an analytical implication that merits explicit acknowledgement. When a composite indicator aggregates dimensions with unequal weights, exceeding the overall threshold may reflect the strength of the more instrumental dimensions (feasibility, 15%; curricular coherence, 20%; collaborative work, 15%) while concealing weaknesses in the dimensions that define the model itself (UDL integration, 25%; pedagogical use of GenAI, 25%). Reporting both the composite score and the dimension-level profile is therefore essential for a transparent interpretation of the findings.

### 4.2. Temporal Fluctuations: The Pedagogical Selectivity Hypothesis

The non-linear trajectory of pedagogical creativity observed across the three PAR phases suggests that the pedagogical appropriation of GenAI did not develop through a simple cumulative process. Instead, the greater variability observed during the final phase points to increasing differentiation in how participants integrated GenAI into their instructional designs. One possible interpretation is that the Final Challenge functioned as a critical threshold, distinguishing predominantly instrumental uses of GenAI from more pedagogically integrated applications grounded in UDL principles (Escala, 2018).

This interpretation should be regarded as a plausible explanatory hypothesis rather than as a demonstrated causal mechanism. The lower average creativity observed during the final phase may also reflect other concurrent factors, including the greater complexity of the Final Challenge, increased time pressure, cumulative fatigue, higher cognitive demands, a deliberate shift towards pedagogical decision-making that relied less on GenAI, or a more demanding application of the Creativity Scale during the final assessment.

Accordingly, the pedagogical selectivity hypothesis offers one interpretation that is consistent with the available evidence without excluding alternative explanations. Future research using finer-grained process data, longitudinal learning traces, classroom observations, and in-depth interviews could help clarify the mechanisms underlying this pattern.

### 4.3. Natural Language AI Tools as Situated Pedagogical Facilitators

The findings suggest that pedagogical creativity was associated with more sophisticated, multimodal, and intentional forms of GenAI appropriation. Rather than indicating that technology itself enhances creativity, the observed pattern highlights the importance of pedagogical mediation in shaping how GenAI contributes to instructional design. The absence of comparable associations during the earlier PAR phases further supports the view that the educational value of GenAI is unlikely to emerge from technology use alone. Instead, its contribution appears to depend on explicit instructional mediation, inclusive pedagogical frameworks, and opportunities for iterative refinement throughout the learning process.

This interpretation is consistent with the differences identified across the pedagogical GenAI use profiles, in which more integrated and pedagogically purposeful patterns of use were associated with higher levels of pedagogical creativity. Given the non-experimental design of the study, these findings should be interpreted as evidence of association rather than as evidence of causal relationships.

### 4.4. Participatory Action Research as a Framework for Contextual Pedagogical Evaluation and Teacher Education

The findings suggest that PAR provides an appropriate framework for the contextual evaluation of pedagogical models in virtual teacher education settings. Beyond supporting iterative refinement, the intervention also functioned as a form of pedagogical modelling, enabling participants to experience the dual role of learners and prospective teachers responsible for designing inclusive learning environments. This dual perspective extends the methodological contribution of the MIIA-R by illustrating how teacher education can integrate pedagogical innovation, reflective practice, and systematic empirical evaluation within a single instructional process.

### 4.5. Relationship to Previous Research and the Contribution of the MIIA-R Model

Similarly, the positive association between GenAI tool diversity and pedagogical creativity aligns with studies highlighting the potential of GenAI to support personalised learning (e.g., Xavier & Dutra, 2026). However, the present study further suggests that

this relationship depends on pedagogical mediation rather than emerging as an automatic consequence of technology use.

With respect to CBL, the findings reinforce previous evidence that authentic challenges promote learner engagement and creativity (e.g., Moreira-Alcivar, 2025; Doulougeri et al., 2024), while adding an important qualification. Authentic public evaluation contexts appear to accentuate differences in participants' pedagogical appropriation of GenAI, a pattern that is compatible with theories of cognitive transformation (Escala, 2018; Leijon et al., 2022; Gallagher & Savage, 2023).

Taken together, these findings position the MIIA-R as more than a synthesis of previously independent approaches. Its principal contribution lies in providing contextually grounded evidence regarding the pedagogical conditions under which UDL, CBL, and GenAI can be coherently integrated within teacher education. Rather than proposing a new learning theory, the model offers a transferable pedagogical architecture that operationalises the interaction among these three approaches through an iterative process of inclusive instructional design, authentic challenge-based learning, and pedagogically mediated GenAI use.

Beyond its pedagogical and methodological contributions, the MIIA-R should also be read in dialogue with the theoretical tradition of adult education. The participants in this study were in-service teachers engaged in postgraduate education, that is, adult learners whose professional experience constitutes both a learning resource and a frame of reference open to critical revision (Knowles et al., 2015; Kolb, 2015). In this regard, the selective and deliberate use of GenAI tools documented in the final PAR phase (what we have termed pedagogical selectivity) can be interpreted not merely as an optimisation of effort, but as an indicator of transformative learning: teams progressively revised their initial assumptions about what technology can and cannot do in instructional design, moving from uncritical adoption to a negotiated, critical integration of AI outputs (Mezirow, 1991; Taylor & Cranton, 2012).

Recent studies such as B. Dang et al. (2025) show that collaboration between students and AI agents in mixed-reality environments involves complex cognitive and socio-emotional interactions, which strengthens the value of designing inclusive learning experiences in which generative AI is integrated under explicit pedagogical criteria rather than as a mere productivity tool (Feng, 2025).

This reading is consistent with recent evidence showing that technology-mediated reflection supports shifts from descriptive to critical reflection in teacher professional development (Li & Walsh, 2023; van Rensburg & Matahela, 2025), and with current analyses warning that the educational value of GenAI depends precisely on the critical consciousness with which it is appropriated by educators (Costa & Murphy, 2025; Lehtinen et al., 2026). The model's contribution thus extends beyond instructional design: it creates the conditions under which postgraduate teacher education can operate as transformative adult learning, rather than as procedural skills training (Benade, 2015).

Overall, the findings support a context-sensitive interpretation of the MIIA-R. Rather than demonstrating that GenAI inherently improves teaching and learning, the study indicates that meaningful pedagogical integration depends on explicit instructional scaffolding, favourable contextual conditions, and participants' progressive pedagogical appropriation of GenAI. Within these conditions, the MIIA-R provides a coherent framework for supporting inclusive and pedagogically purposeful technology integration in teacher education.

## 5. Conclusions

This study presented the design, implementation, and contextual evaluation of the MIIA-R, providing triangulated evidence of its relevance for teacher education in virtual

learning environments. Its principal contribution lies in the coherent integration of UDL, CBL, and GenAI within a unified pedagogical architecture that addresses contemporary demands for inclusive education, authentic learning, and purposeful technological mediation.

The findings indicate that implementing the MIIA-R was associated with sustained participant engagement and the development of feasible and curricularly coherent teaching and learning proposals. At the same time, the results provide only partial support for the pedagogical integration achieved, as the incorporation of UDL remained close to the predefined functional threshold and the pedagogical use of GenAI did not reach it. These findings suggest that meaningful pedagogical appropriation of GenAI requires explicit instructional scaffolding rather than technological access alone.

Hypothesis 1 received only partial and limited support. The MIIA-R enabled students to produce feasible and curricularly coherent instructional proposals, but it did not consistently achieve its central aim: the deep pedagogical integration of UDL and GenAI. This limitation is inherent to the model as implemented and should be addressed in future iterations through more explicit inclusive scaffolding and structured pedagogical mediation of GenAI.

Pedagogical creativity followed a non-linear trajectory, indicating that its development was associated with processes of pedagogical appropriation rather than with technology use alone. Within this framework, GenAI functioned as a situated pedagogical facilitator whose educational value depended on its integration within inclusive pedagogical practices and deliberate instructional mediation.

The study contributes to the field at three complementary levels. Conceptually, it provides contextually grounded evidence supporting the coherent integration of UDL, CBL, and GenAI within a transferable pedagogical architecture. Methodologically, it demonstrates the value of combining Participatory Action Research with concurrent mixed-methods triangulation to evaluate pedagogical innovation in authentic educational settings. Practically, it offers an adaptable framework to support the design of inclusive, technology-mediated teacher education while recognising the importance of contextual adaptation.

Overall, the findings support a context-sensitive interpretation of the MIIA-R. Rather than positioning GenAI as an autonomous driver of educational innovation, the study indicates that its educational value emerges through pedagogical mediation, inclusive instructional design, and sustained opportunities for teacher reflection and pedagogical decision-making.

*Limitations and Future Research*

The findings of this study should be interpreted in light of several limitations. First, the non-experimental, single-group design, based on a convenience sample, precludes causal inference and limits the statistical generalisability of the findings. In addition, the absence of a homogeneous pretest–posttest design across all study variables means that changes in participants' performance were reconstructed through complementary sources of evidence rather than through repeated equivalent measurements.

Second, the initial evaluation of the 80 final teaching and learning proposals and the 178 learning artefacts was conducted by members of the teaching team, which entails a risk of confirmation bias inherent in participatory designs. To address this limitation, an independent rater external to the teaching team and blinded to the study hypotheses re-evaluated a stratified subsample (31.3% of the proposals and 30.3% of the artefacts), yielding good agreement for the overall rubric score (ICC = 0.84) and substantial agreement for the ordinal levels (weighted $\kappa = 0.74$–$0.78$). A residual limitation nevertheless remains, as agreement was estimated on a stratified subsample rather than on the complete set of evidence, and the lowest coefficients were obtained precisely in the two most interpretative

dimensions (UDL integration and pedagogical use of GenAI). Future applications should extend independent double assessment to the full corpus of evidence.

Third, the non-parametric analyses were complemented by pairwise post hoc comparisons using Dunn's test with Bonferroni correction, together with medians and interquartile ranges for each pedagogical GenAI use profile. These procedures made it possible to identify which profiles actually differed from one another rather than reporting an undifferentiated overall effect. However, the reduced size of some subgroups (n = 8 to n = 14) limits the statistical power of specific comparisons, and profile-specific effect sizes were not estimated. These results should therefore be interpreted with caution and replicated in larger samples.

A further limitation concerns the possibility of a Hawthorne effect, as participants were aware that they were taking part in an innovative, closely supported educational experience. This awareness may have influenced their engagement, perceptions of learning, or evaluation of the intervention. Accordingly, the findings should be interpreted within their specific educational context.

Taken together, these limitations identify several priorities for future research. Comparative studies incorporating control or comparison groups, longitudinal designs examining the sustainability of learning outcomes, further psychometric validation of the Creativity Scale, and applications across different educational levels, disciplinary contexts, and teacher education settings would provide additional evidence regarding the transferability and contextual applicability of the MIIA-R across diverse educational settings.

In addition, 42 of the 178 artefacts could not be included in the correlational analysis because they lacked a valid score for at least one of the two variables examined. Although these artefacts remained available for the applicable descriptive analyses, the missing-data mechanism was not formally assessed. Consequently, some degree of selection bias cannot be ruled out, and the correlation between pedagogical creativity and GenAI tool diversity should be interpreted cautiously.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci16091452/s1, Section S1, Creativity Scale in the Use of GenAI; Section S2, Group Reflective Journal Script (Padlet); Section S3, Analytical Rubric for Evaluation of Didactic Proposals; Section S4, Final Satisfaction Survey—Complete Instrument; Section S5, Expert Validation of Assessment Instruments; Section S6, Extended Protocol for Validation of the MIIA-R Model (including Tables S1–S24 and Figures S1–S14).

**Author Contributions:** Conceptualization, A.P.A.; methodology, A.P.A.; validation, A.P.A. and O.E.-M.; formal analysis, A.P.A.; investigation, A.P.A.; resources, A.P.A.; data curation, A.P.A.; writing—original draft preparation, A.P.A.; writing—review and editing, A.P.A. and O.E.-M.; visualization, A.P.A.; supervision, A.P.A.; project administration, A.P.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and received a favourable ethics assessment from the Research Ethics Committee of Universidad Internacional de La Rioja (UNIR) (Protocol PI:200/2025; provisional approval granted on 23 January 2026; final favourable assessment dated 14 April 2026; definitive certificate signed on 21 April 2026).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to ethical and confidentiality restrictions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| CBL | Challenge-Based Learning |
| GenAI | Generative Artificial Intelligence |
| ICC | Intraclass Correlation Coefficient |
| MIIA-R | Integrated Inclusive AI Challenge Model |
| NPS | Net Promoter Score |
| PAR | Participatory Action Research |
| SD | Standard Deviation |
| UDL | Universal Design for Learning |

## References


Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. [CrossRef]

Bastidas, C. F., Hidalgo, N. E., Pucha, A. G., & Parra, D. M. (2025). Formación docente en aprendizaje basado en retos para-STEM primaria: Análisis comparativo de modelos de desarrollo a través de competencias innovadoras modernas [Teacher education in challenge-based learning for primary STEM education: A comparative analysis of development models through modern innovative competencies]. *ASCE Magazine*, *4*(2), 1207–1224. [CrossRef]

Benade, L. (2015). Teachers' critical reflective practice in the context of twenty-first century learning. *Open Review of Educational Research*, *2*(1), 42–54. [CrossRef]

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. *Qualitative Research*, *15*(2), 219–234. [CrossRef]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. [CrossRef]

CAST. (2024). *Universal design for learning guidelines version 3.0.* Available online: https://udlguidelines.cast.org (accessed on 24 August 2026).

Contreras-Villalobos, T., Baleriola Escudero, E., & López Leiva, V. (2024). Digging and building: How transformative mixed-methods research contributes to explaining and responding to educational exclusion and school dropout. *Methods in Psychology*, *11*, 100170. [CrossRef]

Costa, C., & Murphy, M. (2025). Generative artificial intelligence in education: (What) are we thinking? *Learning, Media and Technology*, 1–12. [CrossRef]

Creswell, J. W., & Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*, 22. [CrossRef]

Dang, B., Huynh, L., Gul, F., Rosé, C., Järvelä, S., & Nguyen, A. (2025). Human–AI collaborative learning in mixed reality: Examining the cognitive and socio-emotional interactions. *British Journal of Educational Technology*, *56*, 2078–2101. [CrossRef]

Dang, D., Alani, N. H. S., Nayani, W., & Erturk, E. (2026). Blended learning design in higher education: A systematic review through TPACK and AI role perspectives (2020–2025). *Education Sciences*, *16*(6), 848. [CrossRef]

Denzin, N. K. (2017). *The research act. A theoretical introduction to sociological methods*. Routledge.

Doulougeri, K., Vermunt, J. D., Bombaerts, G., & Bots, M. (2024). Challenge-based learning implementation in engineering education: A systematic literature review. *Journal of Engineering Education*, *113*(4), 1076–1106. [CrossRef]

Echeverria, B. R., & Otero, L. K. (2025). Inteligencia Artificial Generativa como herramienta pedagógica: Una revisión sistemática sobre su impacto en los procesos de enseñanza-aprendizaje [Generative artificial intelligence as a pedagogical tool: A systematic review of its impact on teaching–learning processes]. *SAGA: Revista Científica Multidisciplinar*, *2*(3), 537–550.

Escala, M. J. (2018). Comprendiendo y promoviendo el aprendizaje transformador [Understanding and promoting transformative learning]. *Ciencia y Educación*, *2*(1), 75–76.

Esquivel-Grados, J., & Reyes-Alvarado, S. (2025). Triangulación en la investigación cualitativa y mixta: Una estrategia para la validez y fiabilidad científica [Triangulation in qualitative and mixed-methods research: A strategy for scientific validity and reliability]. *EDUTECH REVIEW. International Education Technologies Review/Revista Internacional de Tecnologías Educativas*, *10*(2), 13–26.

Feng, S. (2025). Group interaction patterns in generative AI-supported collaborative problem solving: Network analysis of the interactions among students and a GAI chatbot. *British Journal of Educational Technology*, *56*(5), 2125–2145. [CrossRef]

Field, A. P. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Gallagher, S. E., & Savage, T. (2023). Challenge-based learning in higher education: An exploratory literature review. *Teaching in Higher Education*, *28*(6), 1135–1157. [CrossRef]

García, L. (2021). COVID-19 y educación a distancia digital: Preconfinamiento, confinamiento y posconfinamiento [COVID-19 and digital distance education: Pre-lockdown, lockdown, and post-lockdown]. *RIED. Revista Iberoamericana de Educación a Distancia*, *24*(1), 9–32. [CrossRef]

Gómez, M., & Cruz, M. P. (2025). El Diseño Universal para el Aprendizaje como enfoque inclusivo en el objetivo 4 de desarrollo sostenible: Revisión sistemática [Universal Design for Learning as an inclusive approach to Sustainable Development Goal 4: A systematic review]. *Pedagogía Social Revista Interuniversitaria*, *(47)*, 47–67. [CrossRef]

Herr, K., & Anderson, G. L. (2015). *The action research dissertation: A guide for students and faculty*. SAGE Publications.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Inteligencia artificial en la educación: Promesas e implicaciones para la enseñanza y el aprendizaje* [Artificial intelligence in education: Promises and implications for teaching and learning]. Centro para el Rediseño Curricular.

Johnson, L. F., Smith, R. S., Smythe, J. T., & Varon, R. K. (2009). *Challenge-based learning: An approach for our time*. The New Media Consortium.

Kemmis, S., & McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 559–603). SAGE Publications.

Knowles, M. S., Holton, E. F., & Swanson, R. A. (2015). *The adult learner: The definitive classic in adult education and human resource development* (8th ed.). Routledge.

Koehler, M. J., Mishra, P., Akcaoglu, M., & Rosenberg, J. M. (2013). The technological pedagogical content knowledge framework for teachers and teacher educators. In *ICT integrated teacher education: A resource book* (pp. 2–7). Commonwealth Educational Media Centre for Asia (CEMCA).

Kolb, D. A. (2015). *Experiential learning: Experience as the source of learning and development* (2nd ed.). Pearson.

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology*, *71*(5), 353–360. [CrossRef] [PubMed]

Lehtinen, A., Channa, F., Näykki, P., Ahlström, E., & Hiljanen, M. (2026). Generative AI for collaborative learning: Fostering critical thinking in teacher education. *Journal of Computer Assisted Learning*, *42*(3), e70256. [CrossRef]

Leijon, M., Gudmundsson, P., Staaf, P., & Christersson, C. (2022). Challenge based learning in higher education—A systematic literature review. *Innovations in Education and Teaching International*, *59*(5), 609–618. [CrossRef]

Li, Y., & Walsh, S. (2023). Technology-enhanced reflection and teacher development: A student teacher's journey. *RELC Journal*, *54*(2), 356–375. [CrossRef]

López, N., Lara, G., Silverio, M., & Pascual, J. (2026). Artificial intelligence in education: Gaps, trends, and ethical challenges. A literature review (2019–2024). *Comunicar*, *34*(84), 42–56.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.

Moral, S. V., & Moreno-Tallón, F. (2025). Inteligencia artificial y educación inclusiva: Soluciones tecnológicas para una enseñanza accesible (revisión sistemática) [Artificial intelligence and inclusive education: Technological solutions for accessible teaching (systematic review)]. *Digital Education Review*, *47*, 62–77. [CrossRef]

Moreira-Alcivar, E. F. (2025). Aprendizaje basado en retos (ABR) para el fomento del pensamiento creativo y divergente en adolescentes: Diseño, implementación y evaluación en contextos escolares del nivel secundario. *Revista Científica Zambos*, *4*(2), 171–184. [CrossRef]

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). Open University Press. [CrossRef]

Reeve, J. (2024). *Comprender la motivación y la emoción* [Understanding motivation and emotion]. John Wiley & Sons.

Taylor, E. W., & Cranton, P. (Eds.). (2012). *The handbook of transformative learning: Theory, research, and practice*. Jossey-Bass.

van Rensburg, G. H., & Matahela, V. E. (2025). Transformative learning in open distance and e-learning: The critical role of nurse educators. *Discover Education*, *4*, 308. [CrossRef]

Veiga, F., Reeve, J., Veiga, C. M., Wong, Z. Y., Martínez, I., de Carvalho, N. A., & Pereira, A. (2026). Higher education student engagement in learning activities: Clarifying concepts and introducing a short scale. *PLoS ONE*, *21*(2), e0340391. [CrossRef] [PubMed]

Xavier, A. B., & Dutra, E. A. (2026). Inteligencia artificial generativa en la educación: Un análisis bibliométrico [Generative artificial intelligence in education: A bibliometric analysis]. *Educación*, *35*(68), 172–198. [CrossRef]

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*, 39. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.