education sciences

MDPI

Article

Validation of a Questionnaire to Explore Artificial Intelligence Use Among Upper-Secondary and Higher Education Teachers

Rodrigo Narváez, Miguel A. Rivera-Navarro and Luis Medina-Gual *

Department of Educational Innovation, Universidad Iberoamericana, Mexico City 01219, Mexico; investiga.die@ibero.mx (R.N.); miguel.rivera@ibero.mx (M.A.R.-N.)
* Correspondence: luis.gual@ibero.mx

Abstract

Research on generative artificial intelligence (AI) in education has shifted from students to teachers, yet available teacher instruments rest on technology acceptance models, professional knowledge frameworks, or AI literacy, and none articulates dispositional conditions, teaching activities, and contextual consequences within a single functional logic. This study extends to teachers an instrument previously validated with students and grounded in Functional Behavior Analysis and the Interbehavioral Model of Didactic Performance. Participants were 233 Mexican upper-secondary and higher education teachers (mean age 47.22 years, SD = 10.44). The study is strictly psychometric in scope: it reports initial reliability and internal structure evidence, not a profile of teacher practice. Reliability was estimated with Cronbach's alpha and McDonald's omega with bootstrap confidence intervals, and the number of components was determined through Horn's parallel analysis. Alpha coefficients ranged from 0.703 to 0.957 (omega, 0.755 to 0.962). The benefits dimension split into two uncorrelated components, pedagogical benefits ($\alpha = 0.953$) and risks to academic integrity ($\alpha = 0.771$). Domains of didactic performance proved unidimensional (63.0% of variance); the consequences dimension behaved as a formative index rather than a reflective scale; and the dispositional dimension failed to separate individual from contextual variables and required redesign. A second-order analysis yielded two orthogonal axes, favourable pedagogical orientation and risk recognition, on which four quadrants are described for illustrative purposes only. The instrument discriminates across self-reported frequency of AI use. Because components were extracted from Pearson correlations on ordinal items in a single convenience sample, the structure reported here is provisional and requires confirmation with polychoric matrices and robust estimators on an independent sample.

Keywords: artificial intelligence; upper-secondary education; higher education; teaching personnel; questionnaire; instrument validation; educational technology; psychometrics

Check for updates

Academic Editor: Yu-Chun Kuo

Received: 27 August 2026
Revised: 24 September 2026
Accepted: 28 September 2026
Published: 2 October 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

The emergence of large language models in education has produced a body of research that, in its first years, concentrated disproportionately on students. This asymmetry is problematic because it is teachers who mediate the incorporation of these tools into the classroom, define the conditions of their legitimate use, and model their responsible employment. Celik et al. (2022) captured that mediation by organising the opportunities AI offers teachers into three moments of the didactic act: planning, through the identification of learning needs; implementation, through immediate feedback and timely intervention;

and assessment, supported by automated essay scoring. Subsequent evidence confirms this pattern. The systematic reviews by Qian (2025) on pedagogical applications of generative AI in higher education and by Tan et al. (2025) on AI in teaching and teacher professional development document frequent uses for lesson preparation, material generation, and feedback; although both also report automated scoring, they stress the need for human oversight of assessment. Cabellos et al. (2024), with 321 university teachers, further showed that teachers' views of these uses are not homogeneous but are organised around their pedagogical beliefs, with constructivist teachers seeing greater potential in generative AI.

In parallel, the literature documents a persistent tension between positive appraisal of these tools and the concerns they raise. Among the latter, the review by Ogunleye et al. (2024) noted the absence of agreed guidelines and the risks associated with academic integrity, while Sun et al. (2024) systematised the problem of distorted information generated by these models. Shata (2025) provided evidence of particular relevance for the present study: in a survey of 294 full-time faculty members at two US public universities, more than a third opted out of using generative AI, for reasons that included the absence of perceived value, tension with professional identity, and the perception of a threat to human intelligence. That work also documented that non-users associated AI with broader negative social consequences, whereas users linked it to innovation, a contrast that anticipates one of the findings reported below.

Systematic measurement of these factors requires specific instruments. The available instruments for teachers fall into three families. The first rests on technology acceptance models. Cabero-Almenara et al. (2024a, 2024b) applied UTAUT2 to a single sample of 425 university teachers in Ecuador, reported in two articles, and found that those holding constructivist pedagogical beliefs showed greater willingness to incorporate AI. Khlaif et al. (2024) used a UTAUT-based model with 358 university faculty members in the Middle East, focusing on the use of generative AI for student assessment, and Nikolic et al. (2024) systematised this line in a review of teaching academics' attitudes, intentions, and behaviours organised around UTAUT. The second family derives from teacher professional knowledge. Celik (2023) developed the Intelligent-TPACK scale, which incorporates transparency, fairness, accountability, and inclusiveness as the ethical components of the knowledge required to integrate AI tools, and Ning et al. (2024) proposed the AI-TPACK scale, subsequently applied in other contexts (e.g., Cindilli et al., 2026). In Spanish, Saz-Pérez et al. (2024) validated a TPACK adaptation for generative AI environments. The third family addresses AI literacy, with instruments such as the AI Literacy Scale for Teachers (Ning et al., 2025) and the AI literacy scale for teachers developed by Younis (2025).

Instruments with related purposes have been developed in the Ibero-American context. Sigüenza Orellana et al. (2024) validated a questionnaire on teacher perceptions of ChatGPT use in higher education; Cabrera Ramos et al. (2025) designed and validated an instrument on knowledge and perceptions of generative AI for pre-service teachers; Gómez-García et al. (2025) built a 25-item, five-factor measure on responsible AI use among pre-service teachers and pedagogues that explicitly incorporates anxiety and enjoyment; and Villa Sánchez et al. (2026) validated a diagnostic questionnaire on teacher competences for the pedagogical integration of AI.

Despite this accumulation, a problem persists that is not limited to instrument availability: research still offers a fragmented understanding of the conditions under which teachers incorporate AI into their practice. The available instruments allow relevant aspects to be studied, such as acceptance, knowledge, literacy, or attitudes, but leave less explored the way in which these conditions operate jointly in situated teaching practice. Reference frameworks such as the one issued by UNESCO (Miao & Cukurova, 2024) set out the competences expected of teachers, but the critical review by Mikeladze et al. (2024) of

AI competence frameworks for educators concludes that all of them are at an early stage of development, that their competences derive mostly from theory or literature reviews rather than from teachers' practice, and that all still require empirical grounding. Moreover, none of the available instruments is grounded in a functional logic linking antecedents, occasion, and consequences of use. Acceptance models predict the intention to use rather than situated behaviour; knowledge frameworks describe what teachers know rather than the conditions that make their application probable. Galindo-Domínguez et al. (2024), with 445 Spanish teachers, reported high declared willingness to use AI alongside low personal experience with it, a gap the study did not set out to explain. The Mexican case is of interest here for three specific reasons rather than as a generic national setting. It combines rapid and largely unregulated uptake of generative AI in higher education with an almost complete absence of institutional policy, so that the contextual consequences the instrument sets out to register vary widely across institutions and are worth measuring. This is also the setting in which the student version of this instrument was developed and validated, which makes the teacher version directly comparable. It is a Spanish-speaking system for which no validated teacher instrument grounded in a functional logic exists, and instruments developed in other languages would require translation and cultural adaptation before use. Regional evidence is, in any case, scarce: available Mexican studies are descriptive or apply instruments developed elsewhere, as in Jiménez Ramírez et al. (2024), who used the CIIAE questionnaire with 359 university teachers.

This need to understand the phenomenon from the teacher's position continues a research agenda initiated with the student instrument and forms part of the comparative architecture proposed by Chao-Rebolledo and Rivera-Navarro (2024), who applied parallel questionnaires with mirror items to teachers and students at a Mexican university, while Narváez and Medina-Gual (2024) developed and validated an instrument to explore AI use among higher education students on the basis of Functional Behavior Analysis (FBA) and the Interbehavioral Model of Didactic Performance (IMDP). The present study validates the teacher version of that instrument.

FBA departs from a precise methodological assumption: explaining why someone does or refrains from doing something requires reconstructing the contingencies in which that behaviour is embedded, rather than attributing it to stable internal dispositions (Froxán Parga, 2020). Within this approach, dispositional variables are of particular interest; Froxán Parga (2020) characterises them as those "conditions of the organism and the environment that have the property of modifying the value of the elements of a sequence, favouring or hindering the establishment of a given functional relationship" (p. 127). Transferred to teachers, this notion requires asking not only what teachers think about AI, but which conditions make its incorporation more or less probable: training received, prior trajectory with technology, institutional support, peer pressure, and the perception that their role might be replaceable. The IMDP, for its part, disaggregates teaching practice into seven domains of performance: didactic planning, competence exploration, criteria specification, illustration, supervised practice, feedback, and assessment (Silva Victoria et al., 2014). This disaggregation is finer than the planning, implementation, and assessment triad with which the international literature has organised teacher uses of AI and allows precise localisation of the moments of the didactic act in which teachers consider support from these tools appropriate. This articulation allows the instrument to be treated not as an end in itself but as a means to operationalise the stated problem and examine how the teacher's relationship with AI is configured.

The study is therefore strictly instrumental in purpose, and this delimitation should be borne in mind throughout. Its aim is to establish initial evidence of the internal consistency and internal structure of the teacher version of the instrument and to examine whether

the resulting scores behave as the theoretical framework would predict in relation to declared familiarity and frequency of AI use. The study does not aim to profile how Mexican teachers use AI, nor to derive recommendations for professional development or educational intervention; substantive claims of that kind would require a design this validation does not provide and are identified in Section 5 as subsequent work rather than as results.

General question. What reliability and internal structure evidence does the teacher version of the instrument yield, and does that structure behave as predicted in relation to declared AI use? Four specific questions follow: (RQ1) What internal consistency evidence do the dimensions of the instrument present when applied to upper-secondary and higher education teachers? (RQ2) What internal structure underlies each dimension and the instrument as a whole, and what modifications relative to the student version are required to represent the phenomenon among teachers? (RQ3) Do the instrument scores differentiate teachers according to their declared familiarity and frequency of AI use, in the direction the framework predicts? (RQ4) What axes jointly organise the second-order variables, and how are participants distributed on the plane those axes define? RQ4 is answered descriptively: the regions of that plane are a reading device, not an empirical typology.

2. Materials and Methods

2.1. Design

An instrumental, cross-sectional study of descriptive and correlational scope was conducted, aimed at gathering initial reliability and validity evidence based on internal structure and relations with other variables. The procedure took as its starting point the one used in the validation of the student version (Narváez & Medina-Gual, 2024) and added methodological improvements such as the joint reporting of alpha and omega with bootstrap intervals, verification of factorability through the KMO and Bartlett tests, and determination of the number of components through Horn's parallel analysis. The delimitation of regions on the second-order plane was retained for descriptive purposes only; no cluster typology was used as evidence of structure.

2.2. Participants

The sample was assembled through non-probabilistic convenience sampling with chain propagation across inter-institutional academic networks. Participants were teachers from upper-secondary and higher education institutions. Dissemination did reach two institutions outside Mexico, but neither of the two respondents concerned registered acceptance of informed consent, so both were excluded at the cleaning stage, and the analytic sample is entirely Mexican. Because dissemination proceeded by successive forwarding rather than from a closed list, it is not possible to establish a sampling frame or compute a response rate. Because dissemination travelled along a small number of institutional networks, the sample also concentrates most of its cases in five affiliations, which limits any claim about the generalisability of the structure reported. Both features of the procedure are revisited in the Limitations section. The final composition is detailed in Section 3.

2.3. Instrument

The questionnaire was administered through Google Forms and organised into a demographic and profile section and five blocks, preceded by an information screen stating the purpose of the study, the delimitation of the term artificial intelligence to natural language models such as ChatGPT and Gemini, the confidential character of the data, their exclusive use for research purposes, and the estimated response time. Continuation was

conditional on explicit acceptance of these conditions. Table 1 summarises the structure of these six sections, the number and format of their items, and the theoretical referent from which they derive.

Table 1. Structure of the instrument administered to teachers.

Dimension	Items	Format	Theoretical Referent
Demographics and profile	8	Mixed	Sample characterisation
Benefits of AI use	19	5-point Likert	Literature on benefits and risks of AI in education
Dispositional variables	14	5-point Likert	Functional Behavior Analysis (Froxán Parga, 2020)
Domains of didactic performance	15	5-point Likert	Interbehavioral Model of Didactic Performance (Silva Victoria et al., 2014)
Current AI use	5	Mixed	Knowledge, frequency, and purpose of use
Consequences in the workplace	11	Yes/No/Do not know	Literature on institutional effects of AI

The benefits, dispositional variables, and domains of didactic performance dimensions used a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The 48 items of these three dimensions constitute the psychometric core of the instrument. The current AI use dimension includes one item requiring voluntary upload of a conversation with a language model that is not analysed here, so analyses of this dimension draw on four items.

The initial version of the items of the three Likert dimensions was drafted with the support of ChatGPT on the basis of the theoretical referents adopted and the literature reviewed, and was reviewed, edited, and refined by the research team before inclusion; this support is recorded in the instrument specification tables and detailed in Section 2.6.

The benefits dimension expands the 15 items of the student version to 19 and incorporates three items concerning risks to academic integrity and student dependence on these tools. The dispositional dimension moves from 7 to 14 items and distinguishes individual variables of a psychological and biological order from contextual variables of the immediate environment, including social and economic ones. The domains of didactic performance dimension translates the seven IMDP domains into 15 concrete activities of teaching practice. The consequences dimension retains the 11 items of the student version, reworded to inquire about events occurring in the teacher's workplace.

2.4. Procedure

Data collection took place between December 2023 and July 2024. Dissemination relied on inter-institutional academic networks, which explains the institutional heterogeneity of the sample. The form evolved during the application period, so that the earliest applications did not include all the items subsequently incorporated. This circumstance is documented in Section 3.1 and was handled through available-case analysis by dimension.

2.5. Data Analysis

Components were extracted through principal component analysis rather than common factor analysis. The decision follows the procedure of the student version, which allows the two validations to be compared directly, and is appropriate to the purpose pursued here, which is to establish how many dimensions the item sets reduce to and whether that reduction reproduces the theoretical blocks, rather than to estimate a latent measurement model. It is also the fitting choice for the consequences block, which records the occurrence of institutional events and is therefore formative rather than reflective by construction. This limitation is stated explicitly in the Limitations section: principal component analysis does not separate common from unique variance, and applying it to five-point ordinal items through Pearson correlations is an approximation. The confirmatory stage, with exploratory factor analysis on polychoric matrices and robust estimators, is set out in Section 5 as the necessary next step. Processing was carried out in Python 3, using the pandas, NumPy, SciPy, scikit-learn, and factor_analyzer libraries. Data cleaning followed five criteria: verification of informed consent, identification of implausible values in the

age and teaching experience variables, removal of the optional e-mail column to guarantee anonymisation, normalisation of institutional denominations, and detection of invariant response patterns.

Reliability was estimated through Cronbach's alpha and McDonald's omega, with omega as the primary estimate and alpha reported alongside it for comparability with previous studies, following Viladrich et al. (2017), with 95% confidence intervals obtained by bootstrap with 2000 resamples, and with corrected item-rest correlations above 0.30 as the discrimination criterion. Validity based on internal structure was examined through principal component analysis on the correlation matrix, with varimax rotation in multi-component solutions. The number of components was determined through Horn's (1965) parallel analysis with 1000 random matrices and the 95th percentile as the threshold, rather than the eigenvalue-greater-than-one criterion, following the recommendation of Lloret-Segura et al. (2014); the Kaiser criterion is reported for comparative purposes only, as it was the one used in the student version. Factorability of each matrix was verified beforehand through the KMO index and Bartlett's test of sphericity. Applying Pearson correlations to five-point ordinal scales is acknowledged as an approximation; to bound its impact, floor and ceiling effects were examined for every item.

The overall structure of the test was explored through a second principal component analysis on the component scores of the Likert dimensions and the summative consequences index. Each participant's position on the plane defined by the two retained components was examined for descriptive purposes and contrasted with a k-means cluster solution evaluated through the silhouette coefficient. Associations between categorical variables were assessed with the chi-square test and Cramér's V; relations between ordinal variables with Spearman's rho; and between-group comparisons with the Mann–Whitney U test and the rank-biserial correlation as effect size.

2.6. Use of Generative Artificial Intelligence

Generative AI tools were used at two moments of the process. During the original design of the instrument, ChatGPT (GPT-3.5, OpenAI, San Francisco, CA, USA), accessed between 23 and 28 November 2023, supported the generation of a first version of the items of the benefits, dispositional variables, and domains of didactic performance dimensions, which were subsequently reviewed, edited, and validated by the research team against the theoretical referents adopted. During manuscript preparation, AI tools were used to assist with literature search and organisation, database cleaning and analysis, and drafting, editing, and translation of the text. The authors reviewed and verified all analyses, reported figures, and cited references and assume full responsibility for the content and accuracy of the work.

3. Results

3.1. Data Cleaning

The original database contained 241 records. Two cases that did not register acceptance of informed consent were removed. Three further records were identified as form tests rather than participant responses: they were submitted from a single address within a fifteen-minute span, declared ages of one, two, and three years, and gave Docente Nivel Bajo, Docente Nivel Medio, and Docente Nivel Alto as institutions of affiliation; they were excluded. One additional implausible age (19,241) was recoded as missing without deleting the case, since the remaining responses of that participant were valid. Three participants showed invariant responses across all 48 Likert items, a pattern that provides no discriminative information and can artificially inflate internal consistency coefficients; these cases were excluded. The final analytic sample comprised 233 teachers.

Twenty-six participants responded identically to the 19 items of the benefits dimension, which contains three items worded in the opposite direction to the rest. These cases were retained because the same participants did vary their responses across the remaining dimensions, but the pattern suggests the need to interleave reverse-worded items with more visible markers in future versions.

The evolution of the form during the application period produced systematic missingness in nine items: 30 cases in three benefits items and one dispositional item, 28 cases in two additional items, and 12 cases in three further items. These missing data do not reflect participant omission but rather the absence of the item from the form version answered, so they were handled through available-case analysis by dimension. Consequently, the effective sample size varies between 203 and 233 depending on the analysis and is reported explicitly in each table.

Examination of the missingness pattern revealed a confound that warrants explicit declaration. The 30 participants who answered the initial form version come almost entirely from a single institution: 26 of them belong to Universidad Iberoamericana Mexico City. That subgroup also reports a substantially higher number of AI-related institutional events (mean of 5.57 versus 3.50 in the rest of the sample). Because these cases lack items from the benefits and dispositional dimensions, they are entirely excluded from the second-order and profile analyses, which rest on 203 participants. Those analyses therefore do not incorporate the institution with the highest declared density of institutional activity around AI, which probably attenuates the variability observed in that dimension. A further contrast confirms that cases without sex data, which correspond to 28 of the 30 cases from the initial version, do not differ from the rest in perceived benefits (3.64 versus 3.52; $p = 0.592$) or in domains of didactic performance (3.66 versus 3.74; $p = 0.568$), but do differ in the consequences index (5.64 versus 3.51; $p < 0.001$). Missingness is therefore not ignorable with respect to that dimension in particular. It should be stressed that these are not data lost at random but items that were never administered to that subgroup, a distinction that determines which remedies are available and is taken up in Section 3.9.

Finally, the optional e-mail column was removed, and institutional denominations corresponding to the same institution written in different ways were normalised. Normalisation covered spelling variants of the same affiliation, among them three renderings of Universidad Autónoma Metropolitana, three of Instituto Politécnico Nacional and three of Universidad Nacional Autónoma de México, together with the upper-secondary and higher education units of Universidad Autónoma de Guerrero. Three participants submitted the questionnaire twice from the same address; these repeated submissions were retained in the principal analysis, and their effect is examined in Section 3.9. Because only 173 of the 233 participants supplied an address, this check is partial, and the possibility of further undetected repeated submissions cannot be excluded.

3.2. Sample Description

The 233 participants had a mean age of 47.22 years (SD = 10.44; Mdn = 46; range 24 to 74 years; $n = 232$ with valid data) and a mean teaching experience of 16.92 years (SD = 11.24; Mdn = 15; range 0.1 to 49 years; $n = 223$). Among those reporting sex ($n = 205$), 54.6% identified as male and 45.4% as female; the 28 cases without data correspond to the initial form version, which did not include this item.

Regarding teaching profile, 36.1% teach simultaneously at undergraduate and postgraduate levels, 35.6% exclusively at the undergraduate level, 18.0% (42 participants) in upper-secondary education, and 10.3% exclusively at the postgraduate level. The presence of the upper-secondary subgroup explains why the scope of the study is declared as upper-secondary and higher education. A contrast between the two levels showed that

they do not differ in any of the three perception dimensions, namely benefits (3.53 versus 3.54; $p = 0.300$), risks (3.38 versus 3.61; $p = 0.306$), and domains of didactic performance (3.85 versus 3.71; $p = 0.115$), but do differ in the institutional consequences index, where upper-secondary teachers report 2.26 events versus 4.10 in higher education (U = 2216.0; $p < 0.001$; r = 0.448). Equivalence across the perception dimensions justifies keeping both subgroups together in the analyses, with the caveat that the consequences index is not comparable across educational levels without controlling for institutional context. On the self-declared continuum between full dedication to teaching and full dedication to research, the central category comprised 116 participants, with 85 oriented towards teaching and 32 towards research.

After normalisation, participants declared 29 distinct institutional affiliations. Table 2 reports the distribution, grouping as Other the institutions that contribute less than 5% of the sample, and adds the sector to which each belongs. Five institutions concentrate 86.7% of the cases, and the remaining 24 contribute 31 participants between them, including Universidad Nacional Autónoma de México, Instituto Politécnico Nacional, Universidad de Guadalajara, Universidad Anáhuac México, Colegio de Postgraduados, teacher-training colleges, and upper-secondary schools. The distribution is reported here rather than only in the Supplementary Materials because the concentration of the sample in a small number of institutions limits the generalisability of the structure reported, as discussed in the Limitations section. The sector breakdown qualifies that concentration in one respect: although five institutions supply most of the cases, public and private institutions are represented in comparable proportions, so the sample is not confined to one segment of the Mexican system. Table S1 gives the complete breakdown.

Table 2. Distribution of the sample by institution of affiliation and sector.

Institution	Sector	*n*	%
ITESO Guadalajara	Private	76	32.6
Universidad Autónoma Metropolitana	Public	55	23.6
Universidad Autónoma de Guerrero	Public	33	14.2
Universidad Iberoamericana Mexico City	Private	26	11.2
El Colegio de la Frontera Sur	Public	12	5.2
Other institutions (24, each <5%)	Public and private	31	13.3
Total		233	100.0

n = 233. Institutions contributing less than 5% of the sample are grouped as Other; Table S1 gives the complete breakdown, with 25 of those participants in public institutions and 6 in private ones. Sector follows the standard classification of the Mexican system: Public covers federal, state, and autonomous public institutions, including public research centres; Private covers privately funded institutions, among them the two Jesuit universities in the sample. Overall, 125 participants (53.6%) are affiliated with public institutions and 108 (46.4%) with private ones.

By area of disciplinary specialisation, the sample is distributed with relative uniformity across seven fields, with none concentrating more than a fifth of participants: social sciences (19.7%), education (18.9%), engineering and mathematics (15.9%), humanities and arts (11.6%), biological and health sciences (11.6%), economic and administrative sciences (9.0%), and exact or natural sciences (9.0%); the remaining 4.3% declared their affiliation by educational level rather than by disciplinary area. The distribution is also presented in the Supplementary Materials (Figure S1).

3.3. Declared Familiarity and Frequency of Use

Declared knowledge differs markedly between the two models examined. For ChatGPT, the median fell at the midpoint of the scale, with 36.1% of participants declaring they knew it “somewhat”, 22.3% “a little”, 18.0% “much”, 12.4% “not at all”, and 11.2% “a great deal”. Gemini presented a markedly different profile: 60.1% declared no knowledge of it whatsoever, and only 8.1% fell in the two highest levels. Both distributions are detailed in the Supplementary Materials (Figure S2).

As to frequency of use, 27.5% of teachers declared that they have never used these models, 17.2% that used them less than once a month, 15.9% once a month, 23.6% once a week, and 15.9% almost daily. The median corresponds to monthly use (Figure S3).

As expected, a strong association was observed between declared knowledge of ChatGPT and its frequency of use, $\chi^2(16, N = 233) = 175.55$, $p < 0.001$, Cramér's V = 0.434, corroborated through Spearman's correlation coefficient, $\rho = 0.731$, $p < 0.001$. This result replicates the one reported in the student version of the instrument and offers a first indication of internal coherence between independent self-report items.

Unlike the finding with students (Narváez & Medina-Gual, 2024), where no differences were detected between undergraduate and postgraduate levels, segmentation by the educational level at which teaching is exercised did yield significant differences in declared knowledge, $\chi^2(4, N = 233) = 28.91$, $p < 0.001$, V = 0.352, when contrasting upper-secondary with higher education. By contrast, no significant differences were observed in frequency of use, $\chi^2(4, N = 233) = 5.08$, $p = 0.279$. The combination of both results suggests that the gap between educational levels lies in declared familiarity with the technology rather than in its effective incorporation into practice.

Sex showed a weak association with frequency of use, $\chi^2(4, N = 205) = 9.60$, $p = 0.048$, V = 0.216, and none with declared knowledge. Neither age nor years of teaching experience were associated with frequency of use ($\rho = -0.074$, $p = 0.260$ and $\rho = -0.026$, $p = 0.698$, respectively), and age showed only a small negative correlation with knowledge of ChatGPT ($\rho = -0.147$, $p = 0.025$). This last result does not support the presumption that teaching seniority by itself predicts resistance to these tools.

3.4. Reliability Analysis

The four dimensions of the instrument were analysed through classical test theory. Table 3 reports Cronbach's alpha and McDonald's omega coefficients, the percentage of items with discriminations above 0.30, and the range of corrected item-rest correlations.

Table 3. Reliability analysis results by dimension.

Dimension	k	*n*	α [95% CI]	ω [95% CI]	Disc. > 0.30	Disc. Range
Benefits (full)	19	203	0.938	0.948	84.21%	0.284 to 0.769
Benefits (refined)	16	205	0.953 [0.936, 0.964]	0.958 [0.945, 0.969]	100.00%	0.667 to 0.768
Academic risks	3	203	0.771 [0.696, 0.830]	0.869 [0.834, 0.898]	100.00%	0.544 to 0.717
Dispositional variables	14	203	0.703 [0.609, 0.767]	0.755 [0.598, 0.821]	64.29%	0.117 to 0.477
Domains of performance	15	221	0.957 [0.939, 0.968]	0.962 [0.948, 0.972]	100.00%	0.639 to 0.830
Consequences	11	233	0.707 [0.652, 0.750]	0.773 [0.737, 0.801]	54.55%	0.072 to 0.547

k = number of items; α = Cronbach's alpha; ω = McDonald's omega; 95% CI = confidence interval estimated by bootstrap with 2000 resamples; Disc. > 0.30 = percentage of items with corrected item-rest correlation above 0.30. Intervals are not reported for the full version of the benefits dimension because its two-dimensional structure invalidates interpretation of the coefficient. The consequences dimension was coded dichotomously, with 1 indicating the reported occurrence of the event and 0 grouping "No" and "Do not know" responses. Variation in n follows from the available-case treatment described in Section 3.1.

Three dimensions reach fully satisfactory coefficients. Domains of didactic performance present the highest consistency in the instrument, with all items above the discrimination criterion and item-rest correlations no lower than 0.639. The benefits dimension, in its full version, yields an alpha of 0.938 but with three items below the criterion: those referring to student dependence on AI, to the difficulty of detecting breaches of academic integrity, and to the promotion of such breaches. These three items share a feature: they are worded in the opposite direction to the rest of the dimension, since they refer to risks rather than benefits.

When these three items are separated, the benefits dimension rises to an alpha of 0.953 and an omega of 0.958, with all 16 items above the discrimination criterion. The three segregated items, in turn, form a brief scale with acceptable consistency, with an alpha

of 0.771 and an omega of 0.869. This result constitutes the main structural modification relative to the student version of the instrument (Narváez & Medina-Gual, 2024), in which the benefits dimension was unidimensional with 15 items.

The dispositional and consequences dimensions present lower coefficients, both around 0.705. In both cases, the discrimination pattern proves more informative than the global coefficient. In the dispositional dimension, five items fall below the criterion, among them those concerning the probability of replacement of the teaching role (0.117), physical limitations in handling technological tools (0.150), perceived pressure from colleagues or the institution (0.167), and the influence of economic availability (0.278). These are precisely the conditions that FBA classifies as dispositional variables of a biological and contextual order, different in nature from the individual psychological variables that dominate the rest of the dimension. In the consequences dimension, five items fall below the criterion, among them the one concerning attempts to block AI access on the institutional network (0.072), the least frequent event in the whole block.

Confidence intervals qualify these readings. In the refined benefits and domains of performance dimensions, the intervals are narrow, and their lower bounds remain above 0.93, so estimation is precise. By contrast, the alpha interval for the dispositional dimension spans from 0.609 to 0.767, and the omega interval from 0.598 to 0.821. With that width, the claim that this dimension attains acceptable consistency cannot be firmly sustained: the lower bound falls below any conventional threshold. The defensible conclusion is more modest, namely that the dispositional dimension, as currently constructed, does not function as a homogeneous scale, which is consistent with the multicomponent structure reported below.

Examination of floor and ceiling effects supports this differential reading. In the benefits, risks, and domains of performance dimensions, no item concentrates more than 28% of responses at one end of the scale, indicating a reasonably dispersed distribution. The dispositional dimension behaves differently: the item on physical limitations gathers 57.5% of responses at the "strongly disagree" end, and the one on the value of expert training concentrates 51.5% at the opposite end. Both are items with restricted variance, which contributes to explaining their low correlations with the rest of the dimension.

3.5. Validity Based on Internal Structure

Before proceeding to principal component analysis, the factorability of each correlation matrix was verified. Table 4 summarises the results by dimension.

All KMO indices reach the 0.70 threshold, with values above 0.94 for the benefits and domains of didactic performance, well above the 0.80 level that Lloret-Segura et al. (2014) regard as satisfactory, and all Bartlett's sphericity tests were significant.

The comparison between retention criteria is informative in itself. The two criteria coincide only for the domains of performance dimension and diverge in the other three: in benefits, the Kaiser criterion retains three components and parallel analysis two; in the dispositional dimension, four against three; and in consequences, four against two. In every case, the divergence occurs in components whose empirical eigenvalues barely exceed unity but fall below the 95th percentile of the distribution generated with random data. For benefits, the third empirical eigenvalue is 1.013 against a random threshold of 1.456; for the dispositional dimension, the fourth is 1.244 against 1.254; and for consequences, the third is 1.210 against 1.240. This is precisely the overextraction scenario that Lloret-Segura et al. (2014) warn against. Subsequent analyses adopt retention by parallel analysis, and in each case what is gained and lost relative to the more liberal solution is indicated. Table S2 reports the full series of empirical eigenvalues and random thresholds for every dimension, so that readers can appraise the retention decisions directly.

Table 4. Sampling adequacy, sphericity, and component structure by dimension.

Dimension	k	*n*	KMO	Bartlett	Horn	Kaiser	1st Var.	Total Var.
Benefits	19	203	0.942	$\chi^2(171) = 2585.91$ ***	2	3	50.7%	61.3%
Dispositional variables	14	203	0.700	$\chi^2(91) = 704.70$ ***	3	4	24.0%	49.2%
Domains of performance	15	221	0.957	$\chi^2(105) = 2578.08$ ***	1	1	63.0%	63.0%
Consequences	11	233	0.707	$\chi^2(55) = 411.68$ ***	2	4	26.0%	38.6%

KMO = Kaiser–Meyer–Olkin sampling adequacy index; Horn = number of components retained by parallel analysis with 1000 random matrices and the 95th percentile; Kaiser = number of components with eigenvalue above one, the criterion used in the student version and reported here for comparative purposes only; 1st Var. = percentage of variance explained by the first component; Total Var. = cumulative percentage explained by the components retained under Horn's criterion. *** $p < 0.001$. Solutions with more than one component were subjected to varimax rotation.

The domains of didactic performance present the clearest structure in the instrument: a single component explaining 63.0% of the variance, with subsequent eigenvalues well below unity. This result replicates the one obtained in the student version, where the uses dimension was also unidimensional, with 73.4% of the variance explained. The theoretical implication deserves attention: teachers do not differentiate among the seven domains of didactic performance when appraising the appropriateness of AI support, but issue a global judgement on the usefulness of these tools for their practice. The IMDP therefore functions as a content coverage criterion for the instrument rather than as an empirically recoverable factor structure.

The benefits dimension extracted two components explaining 61.3% of the variance. Table 5 presents the rotated loadings.

Table 5. Rotated component loadings of the benefits dimension.

Item	C1	C2
Promotes student autonomy in the learning process	0.816	
Fosters creativity and problem solving	0.810	
Has the potential to identify and address individual needs	0.794	
Contributes to creating more interactive educational experiences	0.784	
Can improve the quality and relevance of educational content	0.777	
Can flexibly adapt to different ways of learning	0.776	
Allows assessment methods to be personalised and diversified	0.774	
Increases the efficiency of time devoted to administrative tasks	0.766	0.249
Can provide more detailed and precise feedback	0.761	0.209
Helps identify and prioritise areas of support for students	0.749	
Contributes to developing information management skills	0.748	0.252
Positively influences equality of learning opportunities	0.744	
Facilitates access to diverse and up-to-date information	0.741	
Should be implemented throughout the school curriculum	0.734	
Fosters the development of critical thinking	0.734	
Improves student motivation to participate	0.696	
Could promote breaches of academic integrity		0.784
Makes it harder for teachers to detect integrity breaches		0.782
Could make students overly dependent on these tools		0.877

$n = 203$. Varimax rotation. C1 = perceived pedagogical benefits (50.7% of variance); C2 = risks to academic integrity and autonomy (10.6%). Loadings below 0.20 are omitted for readability. Items are presented in abbreviated form.

The sixteen items referring to benefits load consistently on the first component, with values between 0.696 and 0.816, while the three items referring to risks load exclusively on the second, with values between 0.782 and 0.877. The Spearman correlation between the mean scores of both subsets is 0.109 and does not reach statistical significance ($p = 0.120$), indicating that these are two separable judgements: recognition of pedagogical benefits and perception of risks to academic integrity are not opposed to each other but coexist without a systematic relationship.

The dispositional variables dimension presented the most complex structure, with three components retained by parallel analysis explaining 49.2% of the variance. Table 6 presents the rotated loadings.

Table 6. Rotated component loadings of the dispositional variables dimension.

Item	D1	D2	D3
Implementing AI seems exciting or interesting to me	0.808		
Training and expert modelling would improve my use of AI	0.787		
I understand that I am responsible for the uses I give AI	0.759		
The practice of colleagues using AI motivates me	0.592	0.304	
My institution supports the implementation of AI	0.559	0.224	
I feel prepared to promote ethical use among students	0.465		0.565
I have prior experience using AI in education	0.203	0.766	
I have received specific training on the use of AI			0.767
I experience physical limitations in handling these tools		0.411	
I experience pressure to implement AI		0.749	
It is likely that AI will replace the teaching role			0.637
My past experiences with AI affect my current disposition		0.238	0.302
My past experiences with ICT affect my current disposition	−0.329	0.410	0.384
Economic availability influences the adoption of AI	−0.359	0.520	0.342

n = 203. Varimax rotation on the three-component solution retained by parallel analysis. D1 = favourable disposition and perceived support (24.0%); D2 = history of experiences and perceived barriers (15.4%); D3 = specific training and preparedness (9.8%). The composition of D2 and D3 is not stable across the sensitivity analyses reported in Section 3.9; see that section and the Limitations section. Loadings below 0.20 are omitted.

The three-component solution is interpretable but not entirely clean. The first component clearly groups the conditions that favour the establishment of the functional relationship: declared interest in incorporating AI (0.808), the value attributed to expert training (0.787), the assumption of responsibility for one's own uses (0.759), the modelling influence of peers (0.592) and perceived institutional support (0.559). The second gathers prior experience of using AI in education (0.766) alongside perceived pressure to implement it (0.749), the influence of economic availability (0.520), physical limitations in handling these tools (0.411) and the weight of past experiences with ICT (0.410), so that trajectory and barriers are not separated. The third groups having received specific training (0.767), the perception that AI may replace the teaching role (0.637) and feeling prepared to promote ethical use (0.565), a combination that the label "specific training and preparedness" captures only in part.

Two features qualify the interpretation of this solution. The first is that it is where the two retention criteria differ: under the Kaiser criterion, a fourth component is extracted and the mixture in the second partly dissolves, but parallel analysis indicates that this fourth component does not differ from what would be expected by chance (empirical eigenvalue of 1.244 against a threshold of 1.254). The second, and more consequential, is that the composition of the second and third components is not stable. When the three records later identified as form tests are retained, the history of experiences with technology emerges as a clean third component and training mixes with barriers; with those records removed, trajectory and barriers combine while training separates. That three cases out of 236 should reorganise two components is itself the finding: the first component is robust, and the remaining two are not identified with the precision the theoretical framework would require. Faced with this, the criterion adopted is to report the conservative solution as the result and to treat any reading of D2 and D3 as a structural hypothesis pending confirmation. The data do not support the view that resources and barriers constitute a single factor, but rather that the dispositional block, in its current composition, does not allow the individual psychological variables and the biological and contextual variables of Froxán Parga's (2020) classification to be separated. This is the aspect of the instrument that most urgently requires redesign.

The consequences dimension extracted two components explaining 38.6% of the variance. Table 7 presents the rotated loadings.

Table 7. Rotated component loadings of the consequences dimension.

Item	K1	K2
I know of teachers who routinely integrate AI into their classes	0.763	
I know of teachers who have modelled AI use in their classes	0.723	
There have been courses on AI for teachers at my centre	0.624	
There have been courses on AI for students	0.513	
Teachers regularly use AI for their teaching	0.424	
There is a policy on AI use at the centre	0.402	
Teachers have voiced concern about student use	0.380	0.685
A student has been sanctioned for AI use	0.263	0.643
AI use concerns teachers and authorities at the centre		0.625
A student has used AI in an unreflective manner		0.665
There has been an attempt to block AI on the centre network		0.327

$n = 233$. Varimax rotation on the two-component solution retained by parallel analysis, applied to the dichotomous coding. K1 = integration and training provision (26.0%); K2 = concern and punitive response (12.6%). Loadings below 0.20 are omitted. Internal consistency of the subsets defined by the dominant loading is low ($\alpha = 0.645$ and $\alpha = 0.609$, respectively). These components are reported as a description of how the events co-occur, not as subscales: the block is treated throughout as a formative index.

The internal consistency of the subsets they define is low, with alphas of 0.645 and 0.609. The conclusion is unequivocal and reinforces the reading already advanced: this block does not behave as a reflective scale. The eleven items record heterogeneous institutional facts that can occur independently, as their structure confirms: a centre may have offered courses without having issued a policy or may have sanctioned a student without its teaching staff integrating AI into instruction.

In psychometric terms, this is therefore a formative index rather than a scale. This characterisation corrects an ambiguity in the student version of the instrument (Narváez & Medina-Gual, 2024), where the consequences dimension was analysed as a scale despite obtaining the lowest coefficient and the smallest percentage of acceptable discriminations. Consequently, in subsequent analyses this dimension is represented by a summative index of the number of reported events, whose mean value was 3.77 events (SD = 2.36; Mdn = 4; observed range 0 to 9 out of a possible maximum of 11).

The coding of this index requires justification, since "No" and "Do not know" responses are grouped into a single value, thereby equating absence of the event with participant unawareness. The justification is definitional rather than pragmatic: the item asks the respondent to report events at their own workplace, so what the index measures is institutional activity around AI as it is known to the teacher, and a teacher who is unaware of an event does not report it. Read this way, the high proportion of unawareness documented below is not measurement noise but part of what the index registers. The index is accordingly designated as an index of known institutional AI activity. Two alternative codings were nevertheless examined and are reported in full in Section 3.9: a three-level ordinal scale placing unawareness in an intermediate position, and a proportion over answered items that excludes unawareness altogether. The three codings correlate between 0.77 and 0.96, and the association with declared frequency of use remains positive and significant under all of them ($\rho = 0.310$, 0.218 and 0.185; all $p < 0.005$), so no substantive conclusion depends on the choice. The dichotomous coding is therefore retained, with the caveat that the index must be interpreted as a measure of known rather than verified occurrence.

Figure 1 details the frequency of each event, together with the proportion of teachers who declared not knowing whether it had occurred.

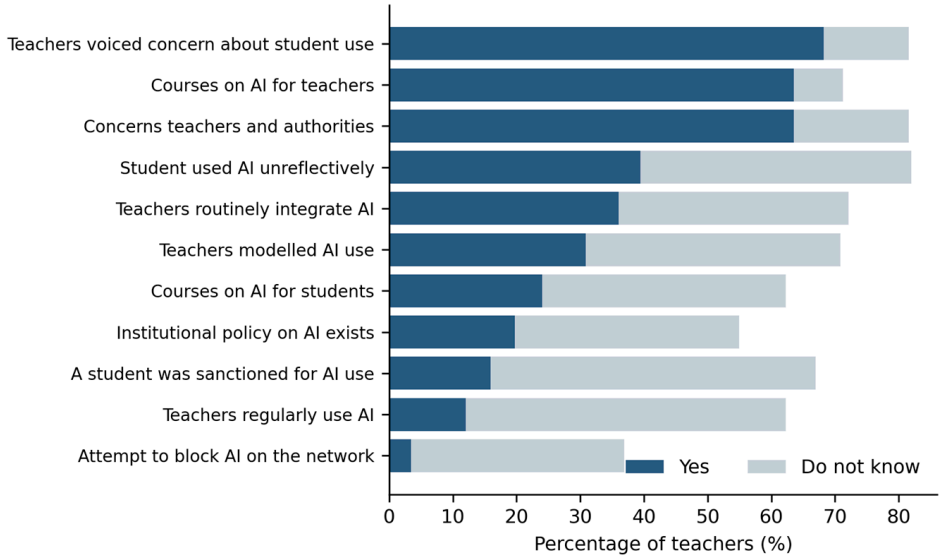


Figure 1. Declared occurrence of AI-related events in the workplace. $n = 233$. The light bar represents the proportion of teachers who answered "Do not know". The complement to 100% corresponds to negative responses.

Three events exceed 60% in terms of declared occurrence: concern voiced by teachers about student use of AI, concern shared by teachers and authorities, and the provision of courses on AI for teaching staff. At the opposite end, only 12.0% state that teachers regularly use AI in their teaching and 3.4% report attempts to block its access from the institutional network. The contrast is telling: AI is established as an institutional concern and as an object of training, but less clearly as routine teaching practice.

Equally notable is the magnitude of unawareness. Half of the participants do not know whether any student has been sanctioned for AI use (51.1%) or whether their colleagues use it regularly (50.2%), and more than a third do not know whether institutional policy on the matter exists (35.2%). This level of opacity about the practices and rules of their own workplace constitutes a finding with direct implications for institutional governance of AI.

3.6. Overall Structure of the Instrument

Following the logic of the student version (Narváez & Medina-Gual, 2024), a second principal component analysis was performed on the component scores of the Likert dimensions and the summative consequences index, that is, on seven second-order variables. The KMO index was 0.607, a mediocre value but above the minimum threshold of 0.50, and Bartlett's sphericity test was significant, $\chi^2(21) = 301.06$, $p < 0.001$. Parallel analysis retained two components explaining 51.0% of the variance; the Kaiser criterion would have retained three, but the third empirical eigenvalue (1.014) falls below the corresponding random threshold (1.131). Table 8 presents the rotated loadings and uniquenesses.

Table 8. Second-order principal component analysis of the instrument variables.

Second-Order Variable	F1	F2	Uniqueness
Pedagogical benefits	0.890	−0.109	0.196
Domains of didactic performance	0.881	−0.067	0.219
Favourable disposition and perceived support	0.789	0.370	0.241
History of experiences and perceived barriers	−0.037	0.487	0.761
Consequences (index)	0.064	0.597	0.639
Academic risks	0.089	0.585	0.650
Specific training and preparedness	0.331	−0.433	0.703

$n = 203$. Varimax rotation. F1 = favourable pedagogical orientation towards AI (33.0%); F2 = risk recognition and contextual exposure (18.1%). Uniqueness expresses the proportion of variance of each variable not explained by the retained components. Variables are grouped by their dominant component.

The first component groups the three central variables of the instrument with loadings between 0.79 and 0.89 and defines an axis of favourable pedagogical orientation towards AI: those who perceive more benefits are also those who recognise more domains of their practice as susceptible to support from these tools and who declare greater disposition and interest. This result converges with the one found in the student version, where the first component grouped perceived benefits, training-related dispositional variables, and perceived uses, although the percentage of variance is lower in the present study (33.0% against 36.1%).

The second component brings together the index of known institutional AI activity (0.597), the perception of academic risks (0.585), and the history of experiences and perceived barriers (0.487), with specific training loading negatively (−0.433). Its interpretation is less immediately evident but coherent: it is an axis of exposure to the phenomenon and recognition of its problematic side. Teachers who acknowledge that their prior trajectory and their working conditions shape their present disposition are also those who register more AI-related events in their workplace and who perceive risks to academic integrity more clearly, while those who report specific training are located at the opposite end. These variables share the feature of reflecting the salience of the phenomenon in the participant's experience, irrespective of whether that experience is appraised positively or negatively. That risk perception loads on this axis and not on the first is informative: recognition of risks is not the opposite of recognition of benefits, but an orthogonal dimension associated with exposure.

The two minor dispositional components present the highest uniquenesses of the set, 0.761 for history of experiences and perceived barriers and 0.703 for specific training and preparedness. That is, they contribute comparatively little to the overall structure of the instrument. Taken together with the instability of their composition documented in Section 3.5, the most prudent reading is that this portion of the dispositional dimension operates with its own logic, largely independent of the rest of the questionnaire, which again points to the need for redesign.

3.7. Teacher Profiles

The two components retained in the second-order analysis define a plane on which each participant's position can be examined. The first axis is interpreted as favourable pedagogical orientation towards AI, with correlations of 0.879 with pedagogical benefits, 0.873 with domains of didactic performance, and 0.809 with disposition. The second axis is interpreted as risk recognition and contextual exposure, with correlations of 0.591 with the index of known institutional AI activity, 0.576 with academic risks, 0.487 with the history of experiences and perceived barriers, and −0.452 with specific training and preparedness. Figure 2 presents the joint representation.

It is worth being explicit about the status of what follows: dividing two continuous axes at their zero point produces four quadrants by construction, so neither the number of quadrants nor their relative sizes constitutes a finding, and they are not teacher types. What follows is a purely descriptive illustration of positions on the plane, retained because it conveys how the two axes combine in practice. The quadrants are named by their position on the two axes rather than by attributes of the teachers located in them, so that the designations remain descriptive of the plane and cannot be read as categories of people. Table 9 presents the results.

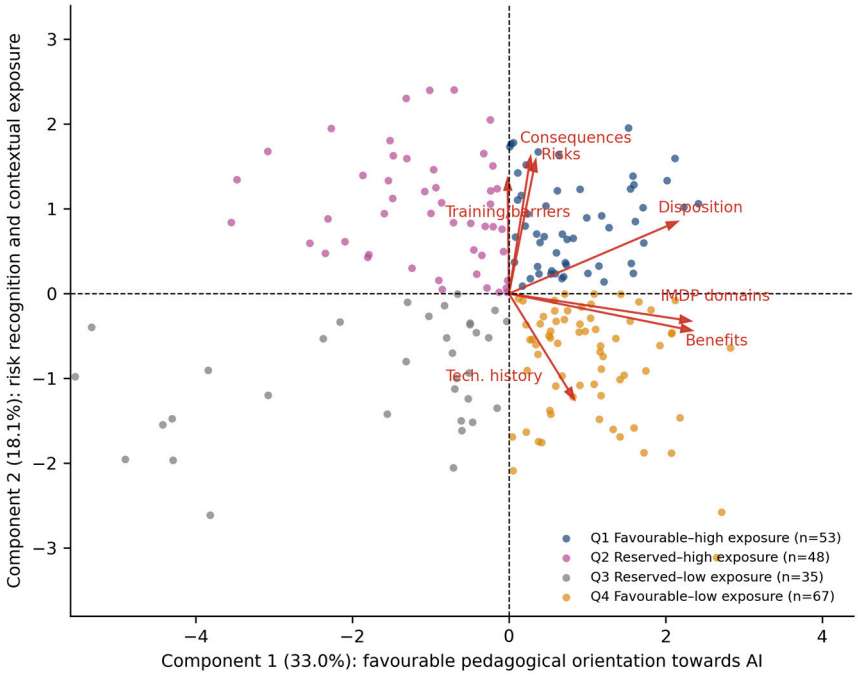


Figure 2. Distribution of participants on the plane of the two second-order components. n = 203. Each point represents a participant. Arrows indicate the contribution of each second-order variable to the two components. Dashed lines mark the origin of each axis and delimit the four quadrants described in the text. The quadrants are a device for reading the plane and do not correspond to empirically distinct groups.

Table 9. Descriptive illustration of the four quadrants of the plane. The quadrants are defined by position on the two axes and are not empirically distinct groups.

Quadrant	*n*	%	Benefits	Risks	Domains	Conseq.	AI Use	% USE
Q1. Favourable–high exposure	53	26.1	3.84	3.91	4.14	5.15	3.25	13.2
Q2. Reserved–high exposure	48	23.6	2.93	3.92	3.10	4.40	1.98	6.2
Q3. Reserved–low exposure	35	17.2	2.80	2.88	3.06	1.26	2.14	40.0
Q4. Favourable–low exposure	67	33.0	4.07	3.40	4.26	2.73	3.24	26.9

n = 203. Benefits, Risks, and Domains express the mean of the items of each subset on a 1 to 5 scale. Conseq. expresses the mean number of events reported in the workplace, out of a maximum of 11. AI Use expresses the mean declared frequency on a scale from 1 (never) to 5 (almost daily). % USE = percentage of upper-secondary education teachers within each quadrant. Quadrant names denote position on the two axes: favourable or reserved pedagogical orientation on the first, high or low risk recognition and contextual exposure on the second.

The four quadrants illustrate how the two axes combine. Q1, favourable–high exposure, gathers positions that pair a high appraisal of benefits with simultaneous recognition of risks and the highest number of known institutional events, together with the most frequent declared use. Q2, reserved–high exposure, gathers positions with a low appraisal of benefits and domains but high risk perception and the lowest frequency of use. Q3, reserved–low exposure, gathers positions with a low appraisal of benefits and domains, low risk recognition, and few institutional events, with an index of 1.26 events, roughly a third of the sample mean. Q4, favourable–low exposure, presents the highest appraisal of benefits and domains in the sample together with comparatively low recognition of risks and of institutional activity. The designations describe where a participant falls on the two continua; they are not four kinds of teacher.

The substantive labels used in the student version of the instrument are not adopted here. Designations such as unaware optimists or disengaged attribute a property to the teachers located in a quadrant, which invites reading the partition as an empirical typology of persons; the analysis supports no such reading. The designations adopted here name

the position on the two axes instead, which is what the quadrant actually identifies, and they are used consistently in the text, in Table 9 and in Figure 2.

A k-means cluster analysis on the same seven second-order variables produces a four-group solution coinciding with this partition in 72.9% of cases, but silhouette coefficients are low across all solutions examined and reach their maximum with two clusters (0.255) rather than four (0.159), indicating that the data do not present a natural grouping structure. The defensible conclusion is that the four quadrants are a useful descriptive device for characterising positions on a two-dimensional continuum, not a typology of discrete categories, and that their value lies in the axes that organise them rather than in the partition itself.

The regions are not independent of the educational level at which teaching is exercised: upper-secondary teachers concentrate in the third and fourth quadrants, where they account for 40.0% and 26.9% of participants, respectively, against 13.2% and 6.2% in the first two. As documented in Section 3.2, this subgroup reports significantly fewer AI-related institutional events, which places its members in the lower half of the second axis even though their perceptions do not differ from those of higher education teachers. Part of the variation captured by that axis is therefore variation in institutional context rather than in individual disposition.

3.8. Relation with Declared Frequency of Use

As additional evidence, the scores of each dimension were contrasted between those who declared using natural language models, for any purpose, and those who declared not using them. The grouping variable comes from an item of the instrument itself, administered in the same session and by the same self-report method, so it does not constitute an external criterion. What these contrasts establish is internal coherence between blocks of the questionnaire that were constructed independently and share no items, which is a necessary but not sufficient condition for speaking of criterion validity. Obtaining an external criterion is set out as a subsequent step. Table 10 presents the results.

All dimensions differentiate significantly between the two groups, with small to moderate effect sizes. Users report a higher appraisal of benefits, greater recognition of domains of practice susceptible to AI support, a more favourable disposition, and a larger number of institutional events registered. The risks dimension operates in the opposite direction: it is non-users who perceive greater risk to academic integrity.

This pattern holds when the relation with declared frequency of use is examined through Spearman correlations, which are positive and significant for benefits ($\rho = 0.263$), domains of didactic performance ($\rho = 0.292$), disposition ($\rho = 0.332$), and the index of known institutional AI activity ($\rho = 0.310$), and negative for risks ($\rho = -0.153$). The convergence between the dichotomous contrast and the ordinal correlation strengthens the interpretation.

Table 10. Contrast between declared users and non-users of natural language models.

Dimension	Users M (SD)	Non-Users M (SD)	U	*p*	r	ρ
Pedagogical benefits	3.64 (0.74)	3.26 (0.98)	6769.5	0.001	0.277	0.263
Academic risks	3.48 (0.93)	3.78 (1.03)	3305.0	0.022	0.206	−0.153
Domains of performance	3.87 (0.69)	3.38 (0.99)	7003.5	<0.001	0.321	0.292
Favourable disposition	0.17 (0.91)	−0.44 (1.10)	5766.0	<0.001	0.386	0.332
Consequences (index)	4.15 (2.28)	2.71 (2.26)	7043.5	<0.001	0.329	0.310

Users $n = 171$, non-users $n = 62$ for the benefits, domains of performance, and consequences dimensions; $n = 146$ and $n = 57$ respectively for risks and disposition, owing to the available-case treatment. U = Mann–Whitney statistic; r = rank-biserial correlation; ρ = Spearman correlation between the dimension and declared frequency of use (n between 203 and 233). Favourable disposition is expressed in standardised component scores.

It is worth noting that the association between declared use and the consequences index invites further interpretation. That users register more AI-related institutional events does not necessarily imply that they work in more active centres, but may indicate that their own practice places them in a better position to perceive what happens around them. The high percentage of "Do not know" responses documented in Section 3.5 supports this interpretation.

3.9. Sensitivity Analyses

Four sensitivity analyses were conducted to appraise how far the results depend on decisions taken during data preparation and on the composition of the sample. All of them are reported in full in the Supplementary Materials (Tables S3–S6).

The first concerns the subgroup whose exclusion from the second-order analyses is documented in Section 3.1. A with-and-without contrast cannot be constructed for the benefits and dispositional dimensions, because the items concerned were never administered to that subgroup: those cases are already absent from the corresponding analyses under available-case treatment, and no amount of imputation can recover responses that were not requested. The comparison is therefore informative only for the two dimensions in which that subgroup does supply data. In domains of didactic performance, excluding it leaves alpha unchanged at 0.957, KMO at 0.957 against 0.956, and explained variance at 63.0%, with a Tucker congruence coefficient of 1.000. In the consequences dimension, alpha moves from 0.707 to 0.703, KMO from 0.707 to 0.675, explained variance from 38.6% to 38.5%, and congruence coefficients are 0.966 and 0.987. Where the comparison can be made, the structure is invariant to the exclusion of that subgroup.

The second compares the participants who answered the initial version of the form with the rest, on the variables both versions share. The two groups do not differ in domains of didactic performance (3.68 versus 3.75; U = 2885.0; $p = 0.643$), but they do differ in the index of known institutional AI activity (5.57 versus 3.50; U = 4603.0; $p < 0.001$; r = 0.512) and, more moderately, in declared frequency of use (3.37 versus 2.75; $p = 0.033$) and in declared knowledge of ChatGPT (3.37 versus 2.87; $p = 0.034$). The pattern is coherent with the confound declared in Section 3.1: the excluded subgroup differs in institutional exposure to AI, not in the perception dimension that both versions share. The consequence for interpretation is stated in the Limitations section.

The third removes the three repeated submissions identified in Section 3.1. Alpha coefficients move by at most 0.007 in any dimension (from 0.953 to 0.954 in refined benefits, from 0.703 to 0.710 in the dispositional dimension, from 0.707 to 0.712 in the consequences dimension, and unchanged in the remaining two), so the independence assumption is not materially compromised by their retention.

The fourth examines the coding of the index of known institutional AI activity, as reported in Section 3.5. Under the dichotomous coding, the index averages 3.77 events (SD = 2.36); placing unawareness at an intermediate value raises it to 5.60 (SD = 2.27), and excluding unawareness from the denominator gives 5.93 (SD = 3.27). The three versions correlate between 0.765 and 0.957, and all three retain a positive and significant association with declared frequency of use ($\rho = 0.310$, 0.218 and 0.185; all $p < 0.005$). Only 31 of the 233 participants did not select the unawareness option for any item, which is itself the reason the third coding cannot serve as the principal one.

4. Discussion

The results support the usefulness of the instrument for exploring the factors associated with AI use by teachers, and at the same time reveal three structural differences relative to its student version, while delimiting what remains to be established.

The first difference concerns the benefits dimension. In the student version, the fifteen items formed a single construct with 68.2% of variance explained. In the teacher version, the incorporation of three items referring to risks to academic integrity produced a clear two-dimensional structure, with practically null cross-loadings and a non-significant correlation between the two subsets. The lack of a statistical association between benefits and risks indicates that teachers do not organise their appraisal of AI along a continuum between enthusiasm and rejection, but through two simultaneous and separable judgements. This observation converges with the literature documenting ambivalent attitudes among university teachers (Cabellos et al., 2024; Khlaif et al., 2024; Nikolic et al., 2024) and, more specifically, with Shata's (2025) finding that those who abstain from using these tools associate AI with broader problematic consequences, while users link it to innovation: in the present study, non-users score higher on risks and lower on benefits, without the two appraisals behaving as the extremes of a single scale. It should be noted that this structure derives from an exploratory analysis on a single sample, so it must be considered plausible rather than established. In practical terms, the recommendation is to report both subscales separately rather than averaging the nineteen items.

The second difference concerns the dispositional dimension. The expansion from seven to fourteen items allowed three components to be recovered, of which only the first, grouping favourable disposition and perceived support, is stable. The second and third, which separate trajectory and barriers from specific training in one solution and combine them differently in another, change composition depending on whether three records later identified as form tests are retained, as documented in Sections 3.5 and 3.9. That instability is not a nuisance to be reported in passing: it is the clearest evidence the study offers that the dispositional block, in its current composition, does not discriminate with the precision that Froxán Parga's (2020) framework would require between individual variables of a psychological order and biological and contextual variables. Any diagnostic use of these two components would be premature. This is the component of the instrument most urgently requiring redesign, and the direction that redesign should take is indicated in Section 5.

Among the dispositional items below the discrimination criterion, those with the lowest values are precisely those inquiring about biological and contextual conditions: physical limitations, external pressure, economic restriction, and perceived replaceability. Two of them also present marked floor or ceiling effects that restrict their variance. Their low discrimination therefore indicates not a substantive deficiency of the content they inquire about, but a poorly discriminating response format and a functioning logic distinct from that of the rest of the dimension, as the high uniqueness in the second-order analysis confirms (0.761 and 0.703). The implication for future versions is twofold: this subset should be treated as an independent scale, and its items require rewording to widen their response range.

The third difference concerns the consequences dimension. The two retained components jointly explain less than 40% of the variance, and the subsets they define present alphas at or below 0.65, indicating that these items do not measure a latent trait but record heterogeneous institutional events. Their appropriate treatment is that of a formative index, and this is how they were handled in the second-order analyses. The examination of alternative codings reported in Sections 3.5 and 3.9 further confirms that the conclusion does not depend on the coding adopted. The dimension is accordingly designated as an index of known institutional AI activity, which states in the name of the construct what the coding already assumed: that the index registers institutional activity as it is known to the teacher, and that unawareness is part of what it measures rather than an artefact of it.

The most robust result of the study is the unidimensionality of the domains of didactic performance, with 63.0% of the variance explained by a single component and all items above 0.639 in discrimination. Teachers do not distinguish among planning, competence exploration, criteria specification, illustration, supervised practice, feedback, and assessment when appraising the appropriateness of AI support. This finding qualifies the planning, implementation, and assessment triad with which Celik et al. (2022) organised the opportunities of AI for teachers, and also the classifications employed by more recent reviews, of pedagogical applications (Qian, 2025) and of AI technologies (Tan et al., 2025): although those organisations are useful for describing the literature, they do not correspond to the way teachers structure their own judgement. The IMDP thus retains its value as a content coverage criterion and as a guarantee that no domain of practice falls outside the instrument, but its seven domains should not be expected to emerge as differentiated factors.

Regarding the quadrants of the second-order plane, the partition into four is a product of the procedure and not a finding; the silhouette coefficients confirm that the data do not present natural groupings, and for that reason the quadrants are designated by their position on the two axes rather than by substantive labels. The substantive contribution is the content of the two axes: among teachers, as among students, favourable orientation towards use and recognition of problematic consequences emerge as orthogonal dimensions. That orthogonality suggests that a more favourable orientation is not, by itself, accompanied by greater risk recognition, or vice versa. It should also be noted that position on the plane is not independent of educational level and that part of the variation on the second axis corresponds to differences in institutional context rather than in individual disposition.

Differences by educational level constitute an unanticipated result. Whereas among students no differences were found between undergraduate and postgraduate levels, among teachers the level at which teaching is exercised is associated with declared knowledge, though not with frequency of use. The dissociation is informative: it suggests that gaps between levels lie in familiarity with the technology rather than in its effective incorporation into practice. In the same vein, no association was observed between years of teaching experience and the use of these tools. This null result, obtained in a convenience sample, does not permit ruling out such an association in the population, but it does withdraw support from the presumption, frequent in institutional discourse, that seniority by itself predicts technological resistance. The finding of Galindo-Domínguez et al. (2024) that the association between digital competence and attitudes towards AI holds regardless of age, sex, or years of experience is consistent with this reading.

Finally, the level of unawareness documented regarding the practices and rules of the workplace itself deserves attention. That half of the teaching staff does not know whether their colleagues use AI regularly or whether any student has been sanctioned for its use, and that more than a third do not know whether institutional policy exists, suggests a deficit of internal communication about AI within institutions, at a time when agreed guidelines for the use of generative AI in higher education are still lacking (Ogunleye et al., 2024). In this sense, the instrument offers an institutional diagnosis that exceeds its original psychometric purpose.

Limitations

The scope of this study is bounded by seven conditions. First, sampling was non-probabilistic by convenience, with chain propagation, so that it is not possible to establish a sampling frame or compute a response rate. The sample is also markedly concentrated: five institutions supply 86.7% of the cases, and a single one supplies almost a third. Because dissemination travelled along institutional networks, participants within each of those five

affiliations share an institutional context, so the effective diversity of settings represented is smaller than the number of participants suggests. The structure reported should therefore be read as evidence obtained in a specific set of Mexican institutions, and its stability across other settings is an open empirical question rather than an assumption. Second, the evolution of the form during the application period generated missing data that are not ignorable with respect to the index of known institutional AI activity, and this excludes from the second-order analyses the institution with the highest declared density of institutional activity around AI. This is a confound between institution and form version, not a random loss: the items were never administered to that subgroup, which is why neither multiple imputation nor a with-and-without contrast is available for the affected dimensions, as set out in Section 3.9. The second-order structure and the plane derived from it were therefore estimated on a subsample with lower institutional exposure to AI, which probably attenuates the variance captured by the second axis. Third, the structures reported derive from exploratory analyses on a single sample, obtained through principal component analysis on Pearson correlations applied to five-point ordinal items, several of which display floor or ceiling effects. Principal component analysis does not separate common from unique variance and, on ordinal data, may misestimate the strength of the underlying structure. The measurement model is consequently not established; its confirmation requires an independent and larger sample, with polychoric matrices and robust estimators. Fourth, within that general limitation, the composition of the second and third dispositional components proved unstable after the removal of three records, so no interpretation of those two components should be carried forward. Fifth, equivalence across educational levels was established through comparison of means rather than measurement invariance analysis, which the size of the smaller subgroup does not permit. Sixth, all evidence of relation with other variables comes from items of the instrument itself, so it constitutes internal coherence rather than criterion validity; obtaining an external criterion to self-report, for example, the effective record of institutional platform use or the observation of teaching practice, is set out as a subsequent step. Seventh, the design is cross-sectional and does not include a retest application, so it does not permit estimating the temporal stability of the scores or establishing directionality; in addition, the data correspond to a specific moment, between December 2023 and July 2024, in which Gemini was widely unknown among teachers, a situation that has very probably changed.

5. Conclusions

This study evaluated a questionnaire to explore the factors associated with artificial intelligence use by upper-secondary and higher education teachers, derived from an instrument previously validated with students and grounded in Functional Behavior Analysis and the Interbehavioral Model of Didactic Performance.

Reliability evidence proved satisfactory in the pedagogical benefits and domains of didactic performance dimensions, with omega coefficients of 0.958 and 0.962 and alpha coefficients of 0.953 and 0.957, respectively, and all their items above the discrimination criterion. The dispositional and consequences dimensions present moderate coefficients, with discrimination patterns that are theoretically interpretable but that, in the dispositional dimension, point to the need for redesign.

Three structural modifications relative to the student version were identified. The first is the splitting of the benefits dimension into pedagogical benefits and academic risks, two separable judgements, with no relevant cross-loadings and no significant correlation between them. The second is the finding that the expanded dispositional block fails to separate individual from contextual variables with precision, which marks it as the

component most urgently requiring redesign. The third is the reconceptualisation of the consequences dimension as a formative index rather than a reflective scale.

The instrument differentiates between declared users and non-users of natural language models across all its dimensions, with small to moderate effect sizes and in the theoretically expected directions. This evidence is one of internal coherence rather than criterion validity, since the grouping variable belongs to the instrument itself.

Finally, the second-order analysis yielded two orthogonal components, favourable pedagogical orientation and risk recognition with contextual exposure. The relevant finding is the orthogonality of these axes, which is consistent with the separation between favourable orientation and consequences observed in the student version of the instrument, and which suggests that a more favourable orientation is not, by itself, accompanied by greater risk recognition.

Subsequent work requires five steps. The first is confirmation of the structure in an independent and larger sample, with estimation through polychoric matrices and robust estimators appropriate for ordered categorical variables, and with formal evaluation of measurement invariance across educational levels and between teachers and students. The second is refinement of the instrument in line with the findings reported: separation of the benefits and risks subscales, redesign of the dispositional block so that individual psychological conditions, prior trajectory, and contextual barriers are written as distinct item sets rather than recovered post hoc, with items reworded to widen their response range, and reformulation of the consequences dimension as an inventory of events with an explicitly handled unawareness option. The third is obtaining validity evidence against a criterion external to self-report. The fourth is the incorporation of a retest application allowing the temporal stability of the scores to be estimated. The fifth is comparative application with the mirror instruments already available for students, which would allow examination of the correspondence between the perceptions of both actors within the same institution, a line of inquiry already initiated in the Mexican context by Chao-Rebolledo and Rivera-Navarro (2024). None of these five steps is dispensable in order to affirm that the instrument is validated; what this study establishes is that the proposed structure is plausible, that three of its four dimensions behave acceptably, and that the fourth requires redesign.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/educsci16101631/s1, Table S1: Distribution of the sample by institution of affiliation; Table S2: Empirical eigenvalues and random thresholds of the parallel analysis for every dimension; Table S3: Structure with and without the subgroup excluded from the second-order analyses; Table S4: Contrast between form versions; Table S5: Effect of removing repeated submissions; Table S6: Alternative codings of the index of known institutional AI activity; Figure S1: Area of disciplinary specialisation of participating teachers; Figure S2: Declared level of knowledge of ChatGPT and Gemini; Figure S3: Declared frequency of use of natural language models.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study. The study was designed and coordinated at Universidad Iberoamericana Mexico City, whose Research Ethics Committee is registered with the National Bioethics Commission (CONBIOÉTICA) of the Mexican Ministry of Health and holds a Federal wide Assurance for the protection of human subjects.

In accordance with that framework and with the Mexican Federal Law on the Protection of Personal Data Held by Private Parties, the study does not fall within the categories of research requiring committee review, because participation was voluntary and conditional on explicit informed consent; participants are adults in professional practice with no relationship of subordination to the research team; the study involves no biomedical intervention and entails no more than minimal risk; and data were anonymised before analysis. Dissemination proceeded through inter-institutional academic networks without the involvement of the administrative bodies of the other participating institutions; respondents participated in a personal capacity, and no institution is an object of evaluation in this work. The study was conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. The first screen of the form stated the purpose of the study, the delimitation of the term artificial intelligence, the confidential character of the data, their exclusive use for research purposes, the optional request for an e-mail address to send an individual profile report, and the estimated response time. Continuation required explicit selection of the agreement option. The two records that did not register this acceptance were excluded from analysis. Participants could omit any item and abandon the questionnaire at any moment without consequence.

Data Availability Statement: The anonymised dataset supporting the results of this study, together with the Python analysis code, is available on reasonable request from the corresponding author. The analysis script reproduces every figure reported in this article from the raw export, and the sensitivity analyses of Section 3.9 are included in it. The version of the dataset made available excludes the optional e-mail column, postal codes, academic unit of affiliation, and free-text fields that could permit the re-identification of participants or institutions. The questionnaire collected one potentially identifiable datum, an optional e-mail address intended for sending an individual report on AI use profile; that field was removed before any analysis and forms no part of the data supporting the results reported here. One item requested the voluntary upload of a conversation with a language model, with the express instruction that it should contain no personal data; that material was not used in the analyses reported here.

Acknowledgments: The authors thank the teaching staff of the participating institutions for their time, and the academic networks that supported the dissemination of the instrument. During the preparation of this manuscript the authors used ChatGPT (OpenAI) for drafting support in the original design of the items, and large language model assistants for literature organisation, data cleaning and analysis support, and for editing and translation of the text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

References

Cabellos, B., de Aldama, C., & Pozo, J.-I. (2024). University teachers' beliefs about the use of generative artificial intelligence for teaching and learning. *Frontiers in Psychology*, *15*, 1468900. [CrossRef] [PubMed]

Cabero-Almenara, J., Palacios-Rodríguez, A., Loaiza-Aguirre, M. I., & Andrade-Abarca, P. (2024a). The impact of pedagogical beliefs on the adoption of generative AI in higher education: Predictive model from UTAUT2. *Frontiers in Artificial Intelligence*, *7*, 1497705. [CrossRef] [PubMed]

Cabero-Almenara, J., Palacios-Rodríguez, A., Loaiza-Aguirre, M. I., & de Rivas-Manzano, M. d. R. (2024b). Acceptance of educational artificial intelligence by teachers and its relationship with some variables and pedagogical beliefs. *Education Sciences*, *14*(7), 740. [CrossRef]

Cabrera Ramos, J. F., Kaechele Obreque, M. T., López Padrón, A., & Álvarez Álvarez, A. (2025). Diseño y validación del cuestionario de conocimiento y percepciones sobre inteligencia artificial generativa para futuros docentes. *Edutec Revista Electrónica de Tecnología Educativa*, *(92)*, 216–233. [CrossRef]

Celik, I. (2023). Towards intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468. [CrossRef]

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, *66*(4), 616–630. [CrossRef]

Chao-Rebolledo, C., & Rivera-Navarro, M. Á. (2024). Usos y percepciones de herramientas de inteligencia artificial en la educación superior en México. *Revista Iberoamericana de Educación*, *95*(1), 57–72. [CrossRef]

Cindilli, B., Sengul, B. Y., Kucuk, S., & Veyis, F. (2026). Exploring gaps and determinants in teachers' artificial intelligence technological pedagogical content knowledge based on a cross-sectional study in Türkiye. *Discover Education*, *5*(1), 402. [CrossRef]

Froxán Parga, M. X. (Ed.). (2020). *Análisis funcional de la conducta humana: Concepto, metodología y aplicaciones*. Pirámide.

Galindo-Domínguez, H., Delgado, N., Campo, L., & Losada, D. (2024). Relationship between teachers' digital competence and attitudes towards artificial intelligence in education. *International Journal of Educational Research*, *126*, 102381. [CrossRef]

Gómez-García, M., Ruiz-Palmero, J., Boumadan-Hamed, M., & Soto-Varela, R. (2025). Percepciones de futuros docentes y pedagogos sobre uso responsable de la IA. Un instrumento de medida. *RIED-Revista Iberoamericana de Educación a Distancia*, *28*(2), 105–130. [CrossRef]

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. [CrossRef] [PubMed]

Jiménez Ramírez, C. R., Martínez Aguirre, E. G., Zárate Depraect, N. E., & Grijalva Verdugo, A. A. (2024). Adopción de la inteligencia artificial en la enseñanza: Perspectivas de docentes de educación superior. *Revista Paraguaya de Educación a Distancia*, *5*(2), 5–16. [CrossRef]

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, *14*(10), 1090. [CrossRef]

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, *30*(3), 1151–1169. [CrossRef]

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. [CrossRef]

Mikeladze, T., Meijer, P. C., & Verhoeff, R. P. (2024). A comprehensive exploration of artificial intelligence competence frameworks for educators: A critical review. *European Journal of Education*, *59*(3), e12663. [CrossRef]

Narváez, R., & Medina-Gual, L. (2024). Validación de un cuestionario para explorar el uso de la IA en estudiantes de educación superior. *Revista Paraguaya de Educación a Distancia*, *5*(4), 29–40. [CrossRef]

Nikolic, S., Wentworth, I., Sheridan, L., Moss, S., Duursma, E., Jones, R. A., Ros, M., & Middleton, R. (2024). A systematic literature review of attitudes, intentions and behaviours of teaching academics pertaining to AI and generative AI (GenAI) in higher education: An analysis of GenAI adoption using the UTAUT framework. *Australasian Journal of Educational Technology*, *40*(6), 56–75. [CrossRef]

Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, *16*(3), 978. [CrossRef]

Ning, Y., Zhang, W., Yao, D., Fang, B., Xu, B., & Wijaya, T. T. (2025). Development and validation of the Artificial Intelligence Literacy Scale for Teachers (AILST). *Education and Information Technologies*, *30*(12), 17769–17803. [CrossRef]

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A systematic review of generative AI for teaching and learning practice. *Education Sciences*, *14*(6), 636. [CrossRef]

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, *69*(5), 1105–1120. [CrossRef]

Saz-Pérez, F., Pizá-Mir, B., & Lizana Carrió, A. (2024). Validación y estructura factorial de un cuestionario TPACK en el contexto de Inteligencia Artificial Generativa (IAG). *Hachetetepé. Revista Científica de Educación y Comunicación*, *(28)*, 1101. [CrossRef]

Shata, A. (2025). "Opting out of AI": Exploring perceptions, reasons, and concerns behind faculty resistance to generative AI. *Frontiers in Communication*, *10*, 1614804. [CrossRef]

Sigüenza Orellana, J., Andrade Cordero, C., & Chitacapa Espinoza, J. (2024). Validación del cuestionario para docentes: Percepción sobre el uso de ChatGPT en la educación superior. *Revista Andina de Educación*, *8*(1), 000816. [CrossRef]

Silva Victoria, H. O., Morales Chávez, G., Pacheco Chávez, V., Camacho Trejo, A. G., Garduño Ortuño, H. M., & Carpio Ramírez, C. A. (2014). Didáctica como conducta: Una propuesta para la descripción de las habilidades de enseñanza. *Revista Mexicana de Análisis de la Conducta*, *40*(3), 32–46. [CrossRef]

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, *11*(1), 1278. [CrossRef]

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, *8*, 100355. [CrossRef]

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, *33*(3), 755–782. [CrossRef]

Villa Sánchez, A., Pizarro Fuentes, E., & Fernandes Queiroz, S. (2026). Validación de un cuestionario para diagnosticar competencias docentes en la integración pedagógica de la inteligencia artificial. *RIED-Revista Iberoamericana de Educación a Distancia*, *29*(2), 153–178. [CrossRef]

Younis, B. (2025). The artificial intelligence literacy (AIL) scale for teachers: A tool for enhancing AI education. *Journal of Digital Learning in Teacher Education*, *41*(1), 37–56. [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.