Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Objective AI Literacy Assessment to Training Needs Mapping in Greek Secondary Education

Nikolaos Nerantzis

nerantzis@uom.edu.gr

University of Macedonia

Anastasia Psalti

University of Macedonia

Kostas Zafiropoulos

University of Macedonia

Research Article

Keywords: AI literacy, secondary education, objective assessment, training needs, teacher professional development

Posted Date: July 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10012281/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence (AI) is increasingly embedded in education, yet evidence about teachers' AI literacy often relies on self-reported readiness rather than objective knowledge. This study mapped objective AI literacy and training needs among Greek secondary education teachers. A cross-sectional online survey was conducted in June 2025 with 271 teachers from mainstream and special-education settings. Participants completed a Greek adaptation of a 26-item objective AI literacy knowledge test for non-technical individuals; responses were scored dichotomously, with incorrect and 'I don't know' responses coded as 0. Total and item-level performance were examined, group differences and partial associations were estimated, and item-level deficits were translated into training priorities using NeedIndex = 1 - $p_{correct}$. Overall performance indicated a moderate baseline (M = 15.82/26, SD = 3.53; overall $p_{correct}$ = 60.85%), but item accuracy varied widely ($p_{correct}$ = 0.09–0.90). The lowest-performing items clustered around foundational technical concepts, including distinguishing automation from AI, differentiating narrow from general AI, understanding the machine-learning workflow, matching AI techniques to applications, and recognizing limits in AI generalization and perception. Scores differed significantly by teaching specialty, ICT training, PC/Internet familiarity, and AI-specific training exposure, whereas demographic differences were limited. NeedIndex mapping identified technical foundations as the highest-priority area for professional development, followed by classroom-relevant interpretation of AI limitations and ethical/governance routines. The study identifies conceptual knowledge patterns that can inform targeted, evidence-informed AI literacy professional development in secondary education.

1. Introduction

1.1 AI in education and the need for objective AI literacy evidence

Artificial intelligence (AI) has moved from a specialized technological field to an everyday infrastructure shaping how information is produced, filtered, and acted upon across social, economic, and institutional life [1–3]. In education, teachers are increasingly expected to evaluate AI-enabled tools, interpret AI-generated outputs, and protect students' learning, privacy, well-being, and fairness. Recent policy and research guidance therefore positions AI literacy as a core professional competence rather than as optional familiarity with emerging tools [4–7].

For teachers, AI literacy should not be reduced to tool use. It involves functional conceptual knowledge about AI mechanisms, data-driven learning, capabilities, limitations, and ethical implications [8, 9]. This distinction is particularly important in the context of generative AI, where classroom discourse may overemphasize prompt use while underemphasizing how AI systems produce uncertain, data-dependent, and sometimes unreliable outputs [4, 10]. Teachers need enough conceptual accuracy to ask better questions about AI systems, evaluate claims about them, and make responsible professional judgments in school contexts.

Despite the growing emphasis on AI literacy, empirical evidence about what teachers actually know remains limited because many studies rely on self-reported readiness, confidence, or attitudes. Self-reports are useful for affective and perceptual constructs, but they may be affected by social desirability, inaccurate self-calibration, and contextual inflation [11, 12]. Objective knowledge assessments can complement self-reports by identifying misconceptions and uneven conceptual coverage that would otherwise remain hidden [13–15].

This measurement issue is consequential in Greek secondary education, where teachers work across diverse disciplinary specialties, mainstream and special-education settings, and uneven professional-learning pathways. Without a localized baseline of objective AI knowledge, professional development risks becoming either too generic to be useful or too technical to be aligned with teachers' actual needs. A baseline study can therefore support evidence-informed capacity building rather than assumption-based training design.

In this context, the distinction between AI awareness and AI literacy is not merely semantic. Awareness may indicate that teachers have encountered AI tools or have heard public debates about generative AI, whereas literacy requires more stable conceptual resources for interpreting AI behaviour. For example, teachers may know that AI can generate text or images, but may still be uncertain about whether these systems understand meaning, why they hallucinate, how training data shape outputs, or why performance may deteriorate when contexts change. Such gaps are professionally relevant because they influence how teachers evaluate AI-supported materials, explain AI to students, and decide when human oversight is necessary.

A localized study is also needed because AI literacy is likely to be shaped by disciplinary opportunity structures. Teachers of ICT or science may have encountered computational or data-related concepts through their disciplinary training, while teachers from humanities, languages, social sciences, arts, or technical specialties may have different entry points and different professional-development needs. This does not imply a hierarchy of capability. Rather, it implies that teacher AI literacy development should be designed with attention to prior exposure, subject-area relevance, and the specific professional decisions that teachers face in secondary schooling.

1.2 Aim, contribution, and research questions

This study examines objective AI literacy among Greek secondary education teachers using a Greek adaptation of the objective AI literacy assessment developed by Ding et al. [13], grounded in the competency-oriented framework of Long and Magerko [8]. The study is positioned as an applied baseline and training-needs mapping study rather than as a comprehensive validation study of the Greek instrument. Its contribution is threefold: it provides an objective baseline, uses item-level diagnostics to locate conceptual fault lines, and translates knowledge deficits into professional-development priorities through NeedIndex = 1 - $p_{correct}$ [16].

This applied orientation is important for the manuscript. The study does not attempt to measure the full ecology of teachers' AI competence, nor does it infer classroom practice directly from test scores. Instead, objective knowledge is treated as a necessary but insufficient layer of professional understanding. By identifying where conceptual knowledge is secure, partial, or fragile, the study provides a basis for designing professional development that is more targeted than general AI awareness training and more accessible than highly technical computer-science instruction.

RQ1. What is the current objective level of AI literacy among Greek secondary education teachers?

RQ2. How does objective AI literacy vary across teachers' demographic, professional, and training-related characteristics?

RQ3. Which teacher characteristics and prior training experiences are associated with higher objective AI literacy performance?

RQ4. Which AI literacy concepts show the largest item-level knowledge deficits?

RQ5. How can item-level knowledge deficits be translated into prioritized professional-development needs for teachers?

2. Conceptual framework

2.1 AI literacy as conceptual knowledge

AI literacy is widely conceptualized as a multidimensional literacy that enables individuals to understand, use, evaluate, and critically engage with AI systems [8, 17, 18]. For non-technical populations, this does not require advanced programming expertise. It requires functional understanding of what AI systems are, how they use data, what tasks they can perform, where their limitations lie, and which social or ethical issues may arise from their use [4, 9].

In this study, AI literacy is treated primarily as objective conceptual knowledge. This includes knowledge of basic AI distinctions, such as automation versus AI, narrow versus general AI, data-driven learning processes, common machine-learning workflows, and typical limitations of AI systems in unfamiliar or ambiguous situations. Such knowledge is not equivalent to pedagogical competence, but it can support teachers' interpretation of AI-generated outputs, recognition of limitations, and responsible professional judgement.

The relationship between technical AI knowledge and pedagogical judgement should therefore be understood as supportive rather than deterministic. A teacher may not know specialized AI terminology and still make careful pedagogical decisions; conversely, a teacher may know technical definitions but use AI tools uncritically. Nevertheless, conceptual knowledge helps teachers formulate appropriate questions: What data might the tool rely on? Is the output probabilistic or deterministic? What human

assumptions are embedded in the system? Could the tool behave differently for different learners or contexts? These questions are central to responsible professional action in AI-enabled schools.

This framing also aligns with broader teacher-knowledge perspectives. From a TPACK perspective, technology-related knowledge becomes educationally meaningful when it interacts with pedagogical and content knowledge [19]. In AI contexts, this interaction also requires ethical and critical dimensions because teachers must evaluate not only whether a tool works, but also whether it is appropriate, fair, transparent enough, privacy-preserving, and educationally justified. Objective AI literacy assessment captures only one part of this professional knowledge ecology, but it can reveal whether teachers possess core concepts needed for deeper pedagogical and ethical reasoning.

2.2 From objective assessment to training-needs mapping

The practical purpose of objective AI literacy assessment in this study is not to rank teachers or certify individual competence. Rather, it is to identify patterns of understanding and misunderstanding that can inform professional-development design. A total score provides a useful baseline, but it can conceal uneven conceptual coverage. Teachers may perform well on familiar ethical or data-related statements while showing weaker performance on foundational technical concepts such as AI system types, machine-learning workflows, or generalization limits.

For this reason, the study emphasizes item-level analysis. NeedIndex, calculated as $1 - p_{correct}$ for each item, transforms item-level deficits into an interpretable prioritization metric: higher values indicate lower item-level performance and stronger potential need for training. This links measurement to professional-development planning and allows training content to be sequenced according to observed conceptual need rather than assumed relevance.

The NeedIndex approach is deliberately simple. It does not replace psychometric validation, nor does it claim that a single item fully represents a complete competence domain. Its value lies in its interpretability for professional-development planning: an item with very low $p_{correct}$ signals a concept that many teachers did not answer correctly and, therefore, a likely candidate for foundational instruction. When several low-performing items cluster around related concepts, the evidence becomes more useful for designing modules, sequencing content, and selecting examples for teacher learning.

This logic is particularly relevant for AI literacy because the construct is broad. A professional-development programme that treats AI literacy as a single theme may overemphasize visible tools or ethical slogans while leaving conceptual mechanisms underdeveloped. Conversely, item-level diagnostics can indicate whether teachers need first to consolidate basic distinctions, then examine classroom cases, and finally apply ethical governance routines. The study therefore uses assessment not as an end in itself, but as an input for targeted capacity building.

3. Methods

3.1 Study design, setting, and participants

A cross-sectional online survey was conducted in June 2025 to assess objective AI literacy among secondary education teachers in Greece and to derive training priorities from item-level performance. Eligible participants were teachers working in lower-secondary and upper-secondary education, including mainstream schools and special-education structures. Recruitment followed a convenience sampling approach through professional and personal networks. The invitation described the study purpose, voluntary participation, anonymity, and the absence of directly identifying data. Because the questionnaire link was distributed through open networks rather than a closed sampling frame, a response rate could not be calculated. After removal of duplicate and incomplete submissions, the final analytic sample comprised 271 teachers.

This recruitment strategy is appropriate for an exploratory baseline and needs-mapping study, but it also requires cautious interpretation. Teachers who respond to an online survey about AI may be more interested in digital technologies, more engaged in professional learning, or more confident with online questionnaires than the wider teacher population. For this reason, descriptive estimates should not be read as nationally representative prevalence estimates. The analytic value of the sample lies primarily in mapping patterns of objective performance, group differences, and item-level priorities within a heterogeneous teacher group.

Table 1
Participant characteristics (N = 271)

Characteristic	n (%)
Gender: Man	66 (24.40%)
Gender: Woman	203 (74.90%)
Gender: Other / I do not want to answer	2 (0.70%)
Highest degree: University	74 (27.30%)
Highest degree: Postgraduate	179 (66.10%)
Highest degree: Doctorate	18 (6.60%)
Age group: 26–35	17 (6.30%)
Age group: 36–45	64 (23.60%)
Age group: 46–55	112 (41.30%)
Age group: 56+	78 (28.80%)
Work area: Urban	174 (64.20%)
Work area: Semi-urban	65 (24.00%)
Work area: Rural	16 (5.90%)
Work area: Island	16 (5.90%)
Specialty: Theology	11 (4.10%)
Specialty: Greek Literature	77 (28.40%)
Specialty: Mathematics	28 (10.30%)
Specialty: Science (Physics, Biology, Chemistry, Geology)	56 (20.70%)
Specialty: ICT	37 (13.70%)
Specialty: Foreign Literature	22 (8.10%)
Specialty: Civic Education	10 (3.70%)
Specialty: Culture & Activities	10 (3.70%)
Specialty: Technical & Technological Direction	20 (7.40%)
Years of service: 0–6	56 (20.70%)
Years of service: 7–12	26 (9.60%)

Note. *Specialty groups were formed to ensure interpretable subject-area comparisons; minimum group size was N > = 10.*

Characteristic	n (%)
Years of service: 13–20	72 (26.60%)
Years of service: 21–30	86 (31.70%)
Years of service: 31+	31 (11.40%)

Note. *Specialty groups were formed to ensure interpretable subject-area comparisons; minimum group size was N > = 10.*

3.2 Instrument, adaptation, and scoring

AI literacy was measured using an objective knowledge test adapted from Ding et al. [13], grounded in the AI literacy framework of Long and Magerko [8]. The Greek version comprised 26 items. The instrument was translated into Greek and back-translated to support conceptual and semantic equivalence, following established guidance for cross-cultural adaptation [20]. Pilot checks focused on item clarity, response flow, and completion time; minor wording adjustments improved readability without altering item meaning or scoring logic.

Each item was scored dichotomously: correct responses received 1 point, while incorrect and 'I don't know' responses received 0 points. The total AI literacy score was computed as the sum of correct responses across the 26 items, yielding a possible range from 0 to 26. Higher scores indicate higher objective conceptual AI knowledge. To translate item performance into professional-development priorities, $p_{correct}$ was computed for each item and NeedIndex was calculated as 1 - $p_{correct}$. NeedIndex was used as a descriptive prioritization metric rather than as a psychometric scale score.

The explicit 'I don't know' option was retained because the study's practical goal was training-needs identification rather than high-stakes certification. In this context, treating 'I don't know' as incorrect produces a conservative estimate of functional knowledge. It also differentiates between concepts that teachers can answer confidently and those that may be uncertain or vulnerable to guessing. This scoring choice is especially relevant for professional development, where uncertainty itself can be informative for designing instruction.

3.3 Statistical analysis and ethics

Analyses addressed overall performance, group differences, correlates and predictors, item-level conceptual gaps, and training-needs prioritization. Descriptive statistics were computed for total scores and item-level $p_{correct}$. Because the total score deviated from normality, group differences were examined using Kruskal-Wallis tests; significant omnibus tests were followed by Dwass-Steel-Critchlow-Fligner post-hoc comparisons. Effect sizes for Kruskal-Wallis tests were reported using epsilon-squared (epsilon2). Associations with ordinal indicators were examined using Spearman correlations. Multiple linear regression estimated partial associations between teacher characteristics, prior training experiences, and total AI literacy score. Primary interpretation of regression coefficients relied on bias-corrected and accelerated bootstrap 95% confidence intervals, with HC1 robust standard errors used as

a sensitivity analysis. Statistical significance was evaluated at alpha = .05, with emphasis on effect sizes, confidence intervals, and practical relevance.

The analytic strategy was therefore descriptive and explanatory rather than predictive. Non-parametric group comparisons were used to respect the observed distributional properties of the total score, whereas regression was used to examine whether training-related variables retained associations with performance after accounting for other covariates. Post-hoc tests and sensitivity analyses are reported in supplementary materials to keep the main text focused on substantive interpretation while preserving transparency.

The study was conducted under institutional ethics approval by the Research Ethics Committee of the University of Macedonia (EHDE, protocol 36/6-6-2025). Participation was voluntary; respondents viewed an information sheet and provided electronic informed consent before accessing the questionnaire. Responses were collected anonymously, and no directly identifying information was stored.

4. Results

4.1 Overall objective AI literacy performance (RQ1)

Across the final analytic sample, total AI literacy scores ranged from 2 to 24 out of 26. The mean score was 15.82 (SD = 3.53), the median was 16.00, and the overall proportion correct was 60.85%. The distribution therefore indicated a moderate overall baseline, but also substantial variation across participants. Score-band classification showed that 3.7% of participants fell into the very low band, 19.9% into the low band, 29.5% into the moderate band, 39.1% into the high band, and 7.7% into the very high band. A Shapiro-Wilk test indicated that the total score deviated from normality, $W = 0.974$, $p < .001$, supporting the use of non-parametric procedures for group comparisons.

Table 2
Summary of total objective AI literacy score

Metric	Value
Items / score range	26 / 0–26
N	271
Observed range	2–24
Mean (SD)	15.82 (3.53)
Median	16.00
Overall $p_{correct}$	60.85%
Shapiro-Wilk test	W = 0.974; p < .001
Score band: Very low	3.7%
Score band: Low	19.9%
Score band: Moderate	29.5%
Score band: High	39.1%
Score band: Very high	7.7%

4.2 Group differences and predictors (RQ2-RQ3)

Objective AI literacy differed significantly by teaching specialty, ICT training level, self-reported PC/Internet familiarity, and AI-specific training exposure. The largest effects were observed for teaching specialty and AI training exposure, suggesting that disciplinary background and structured AI-related learning opportunities were more strongly associated with objective AI literacy than general demographic characteristics. Differences by gender, age group, highest degree, work area, and years of service were not statistically significant.

Bivariate associations showed that total AI literacy score was positively associated with ICT training (rho = .289, p < .001) and PC/Internet familiarity (rho = .196, p = .001), whereas age and years of service were not associated with performance. In the multiple regression model, AI-specific learning experiences were the dominant predictor set. Compared with short-term seminars/workshops, semester or annual programmes, certified knowledge, and undergraduate AI coursework were associated with higher total score estimates, whereas reporting no AI training was associated with lower performance. The model explained a meaningful proportion of variance in total AI literacy performance (adjusted R2 = .204).

Table 3
Key group-difference and predictor results

Analysis focus	Result	Interpretation
Teaching specialty	Kruskal-Wallis: chi2(8) = 43.5, p < .001, epsilon2 = .161	Largest omnibus effect; ICT teachers scored higher than several non-ICT specialties.
AI training exposure	Kruskal-Wallis: chi2(5) = 39.1, p < .001, epsilon2 = .145	Structured AI learning exposure was associated with higher objective performance.
ICT training level	Kruskal-Wallis: chi2(4) = 27.8, p < .001, epsilon2 = .103	Higher ICT training aligned with higher AI literacy scores.
PC/Internet familiarity	Kruskal-Wallis: chi2(2) = 10.4, p = .005, epsilon2 = .039	Digital familiarity showed a smaller but significant association.
Spearman correlations	ICT training: rho = .289, p < .001; PC/Internet familiarity: rho = .196, p = .001	General digital competence may support, but does not replace, AI-specific knowledge.
Regression model	Adjusted R2 = .204	AI-specific training variables formed the most salient predictor set.

Note. *Full omnibus tests, post-hoc comparisons, and HC1 sensitivity results should be provided in Supplementary Tables S1-S3.*

Post-hoc comparisons indicated that the specialty effect was mainly driven by higher performance among ICT teachers compared with several non-ICT groups. This pattern is unsurprising, but it is still important: it suggests that disciplinary proximity to computational and data-related concepts may provide an advantage in objective AI literacy performance. However, the result should not be interpreted as a deficit label for other specialties. Instead, it supports differentiated professional development that offers more explicit technical scaffolding for teachers whose disciplinary background has not routinely included such concepts.

The AI training exposure effect provides a complementary interpretation. Teachers with more sustained or formal AI-related learning experiences generally showed higher performance than those with no AI training. The strongest positive coefficient was observed for undergraduate AI coursework, although this subgroup was small and should be interpreted cautiously. More broadly, the results suggest that structured exposure matters. Short seminars may raise awareness, but longer or certified learning experiences may provide more stable conceptual foundations.

Table 4
Multiple linear regression predicting total AI literacy score: AI training exposure coefficients

Predictor (reference: short-term seminar/workshop)	B	BCa 95% CI low	BCa 95% CI high
Semester / annual programme	1.512	0.222	2.678
Certified knowledge	1.844	0.327	3.172
Undergraduate AI course	3.476	1.117	5.802
Postgraduate AI course	1.821	-0.074	3.725
No AI training	-1.365	-2.398	-0.359

4.3 Item-level conceptual gaps and NeedIndex priorities (RQ4-RQ5)

Item-level performance varied substantially across the 26 items, with $p_{correct}$ values ranging from .092 to .897. The five lowest-performing items concerned whether voice-activated digital assistants are examples of general AI, the basic machine-learning sequence, matching AI techniques to applications, AI recognition of unfamiliar or atypical objects, and whether an automatic washing machine is an AI device. These items indicate pronounced gaps in foundational technical concepts: narrow versus general AI, automation versus AI, machine-learning workflow, technique-to-application mapping, and limits in perception and generalization.

The highest-performing items concerned data interpretation, types of data, the misconception that AI systems are always smarter than humans, AI use in dangerous work, and whether AI always makes fairness decisions. This pattern suggests comparatively stronger performance on selected general, data-related, or ethical items than on foundational technical concepts. NeedIndex mapping confirmed this profile: items with $p_{correct}$ below 40% fell into the very high need band and clustered around technical foundations.

The item-level pattern also helps interpret the meaning of the moderate total score. If only the mean score were considered, the sample might be described as having an adequate or emerging AI literacy baseline. The item profile tells a more nuanced story. Some AI literacy components appear relatively secure, while others are fragile. This unevenness is central for training design because teachers do not need equal attention to every concept. The priority is to identify concepts that are both foundational and widely misunderstood.

In this sense, the lowest-performing items are not isolated test weaknesses. They form a coherent conceptual cluster around AI identification, AI system types, machine learning, and generalization. These concepts underpin later discussions of reliability, bias, privacy, fairness, and accountability. For example, teachers can better understand algorithmic bias when they understand that AI systems learn from data,

and they can better evaluate reliability when they understand that model performance may not generalize across unfamiliar inputs or contexts.

Table 5
Highest- and lowest-performing items

Item	$p_{correct}$
Lowest-performing items	
Q06. Voice-activated digital assistants (Alexa and Siri) are examples of general AI.	0.092
Q15. Put the machine-learning steps in order: collect data, train model, deploy model.	0.295
Q14. Match AI techniques (NLP, reinforcement learning, decision tree) to examples.	0.299
Q08. It is easy for an AI system to recognize objects in unfamiliar or atypical situations.	0.328
Q01. An automatic washing machine is an example of an AI device.	0.358
Highest-performing items	
Q20. Data can be error-prone and requires interpretation.	0.897
Q17. Identify which option is not a type of data.	0.889
Q04. AI systems are always smarter than humans.	0.871
Q11. AI can help humans avoid dangerous work.	0.867
Q24. AI always makes fairness decisions.	0.860

Table 6
Training-needs prioritization bands (NeedIndex = 1 - $p_{correct}$)

NeedIndex	$p_{correct}$	Interpretation
> 60%	< 40%	Very high need: foundational technical concepts such as AI identification, AI system types, machine-learning workflow, and limits of AI perception/generalization.
40–60%	40%-59.9%	High need: partial or unstable understanding; concepts should be addressed through structured modules and applied examples.
20–40%	60%-79.9%	Moderate need: basic competence present, but concepts require applied consolidation and classroom-relevant practice.
< 20%	>= 80%	Low need: comparatively stronger performance; training can focus on transfer, application, and reflective use.

5. Discussion

5.1 Main interpretation

This study mapped objective AI literacy knowledge among Greek secondary education teachers and translated item-level knowledge gaps into professional-development priorities. The aggregate score indicated a moderate baseline, but item-level results showed that this average concealed substantial conceptual unevenness. Teachers performed comparatively well on selected data-related, ethical, or general misconception items, but much less well on foundational technical concepts. This finding supports the argument that general exposure to AI discourse does not necessarily produce stable conceptual understanding of how AI systems work or why they fail [13–15].

The lowest-performing items identify threshold concepts for teacher AI literacy. Distinguishing narrow from general AI, differentiating automation from AI, understanding the machine-learning workflow, and recognizing limits in perception and generalization are not advanced computer-science topics. They are basic conceptual foundations that help teachers evaluate AI-related claims, interpret AI outputs, and understand why issues such as bias, uncertainty, reliability, and accountability arise in AI-enabled educational systems.

This finding has particular significance for secondary education. Teachers are often positioned as mediators between policy discourse, commercial AI tools, student use, parental concerns, and school-level expectations. If their conceptual understanding is uneven, they may struggle to explain AI systems to students, evaluate vendor claims, or distinguish realistic educational affordances from inflated promises. Foundational AI literacy therefore has a protective function: it can reduce susceptibility to hype, support more accurate risk assessment, and strengthen teachers' capacity for human-centred oversight.

5.2 Training exposure, opportunity structures, and the limits of ICT familiarity

Group comparisons and regression analyses suggested that objective AI literacy was more strongly associated with disciplinary and training-related characteristics than with general demographic variables. AI-specific training exposure emerged as the most salient predictor set, while age and years of service were not associated with objective performance. Although the cross-sectional design prevents causal inference, this pattern is consistent with the interpretation that AI literacy is a trainable professional knowledge domain rather than a competence that develops automatically through teaching experience.

At the same time, these findings should be read through the lens of opportunity structures. Teachers enter AI professional development with different disciplinary backgrounds, prior ICT experiences, and access to AI-related learning. ICT training and digital familiarity may support engagement with AI, but they are not equivalent to AI literacy. A digitally fluent teacher may still lack conceptual understanding of how machine-learning systems differ from conventional software, how AI outputs are shaped by data and model design, or why AI systems may fail in unfamiliar contexts. Professional development should therefore be differentiated and staged rather than built on the assumption of a uniform starting point.

The absence of strong demographic effects is also noteworthy. Age and years of service did not explain objective AI literacy performance, suggesting that AI literacy is not simply a generational issue or a by-product of professional seniority. This matters for policy because it challenges common assumptions that younger teachers will automatically be AI literate or that experienced teachers will necessarily lack relevant competence. Instead, structured learning opportunities and disciplinary/professional exposure appear more informative than broad demographic categories.

5.3 From knowledge deficits to professional-development design

The NeedIndex approach provides a practical bridge between assessment and training design. Instead of treating AI literacy as a single undifferentiated construct, item-level analysis identifies which concepts require whole-group foundations, which require targeted modules, and which can be addressed through consolidation and application. In this sample, the first priority is not only ethical awareness or tool demonstration, but conceptual accuracy around what AI is, how AI systems learn, and what limitations they have.

This does not mean that technical knowledge is sufficient for responsible classroom practice. Objective AI literacy knowledge is one layer within a broader professional knowledge ecology that also includes pedagogical content knowledge, ethical reasoning, inclusion, classroom context, and institutional judgement. This interpretation is consistent with TPACK, which frames educational technology use as the integration of technological, pedagogical, and content knowledge [19], and with recent work arguing that teacher AI literacy should combine technical, pedagogical, and ethical-democratic dimensions [21]. The present study therefore identifies conceptual vulnerabilities that may constrain informed interpretation; it does not claim to evaluate teachers' overall AI competence or classroom enactment.

This cautious interpretation also protects the study from an overly narrow technicist reading. The aim is not to argue that teachers must master AI engineering before using AI tools. Rather, the findings show that certain conceptual foundations are needed if teachers are to interpret AI tools responsibly. Pedagogical use should remain grounded in learning goals, inclusion, subject knowledge, and student needs. Technical concepts matter because they inform judgement; they do not replace professional judgement.

6. Implications for teacher professional development

The findings support a tiered professional-development model. Effective teacher learning opportunities should be content-focused, coherent, connected to professional practice, and sustained over time, while also remaining responsive to local needs and avoiding mechanical application of generic professional-development features [22–25]. In the context of AI literacy, this means moving beyond isolated tool demonstrations toward sequenced capacity building.

The proposed sequence begins with foundations because ethical and pedagogical discussions become more powerful when teachers understand the mechanisms that produce AI-related risks. Discussions of fairness, for example, are clearer when teachers understand that systems are trained on data that may be incomplete, biased, or contextually misaligned. Discussions of reliability are clearer when teachers understand generalization limits. Discussions of accountability are clearer when teachers understand that outputs are generated through model-based inference rather than human-like understanding.

Tier 1 should address foundational technical concepts: AI versus automation, narrow versus general AI, the machine-learning workflow, data in model development, and common AI techniques. Tier 2 should focus on AI capabilities, limitations, uncertainty, and failure modes through output-evaluation tasks and classroom-relevant vignettes. Tier 3 should connect conceptual understanding to pedagogical judgement, including feedback, assessment, differentiation, lesson planning, student support, and information evaluation. Tier 4 should address ethics and governance, including privacy, bias, transparency, accountability, human oversight, and school-level routines for responsible use.

The tiers should not be interpreted as rigid or isolated units. In practice, effective professional development can spiral across them: a foundational concept can be introduced briefly, revisited through a classroom vignette, and then connected to ethical or governance questions. For example, a session on the machine-learning pipeline can lead into discussion of biased datasets, automated feedback, student privacy, and the teacher's responsibility to verify outputs before using them in assessment or support decisions.

For inclusive and special-education settings, the staged model also needs explicit attention to accessibility, vulnerability, and human oversight. AI tools may offer support for differentiation, communication, translation, summarization, or feedback, but they may also reproduce assumptions that do not fit students with diverse learning, communication, sensory, or psychosocial needs. Professional development should therefore include cases in which teachers evaluate whether AI-supported recommendations are appropriate for specific student profiles and whether additional safeguards are required.

NeedIndex can also support evaluation of professional-development programmes. Items linked to targeted modules can be administered before and after training to examine whether specific conceptual gaps narrow. However, pre/post item improvement should be interpreted as evidence of conceptual learning, not as direct evidence of classroom implementation. Stronger evaluation designs should combine objective knowledge items with scenario-based tasks, teacher reflections, lesson-design evidence, and classroom-oriented indicators of responsible AI use.

A practical implementation pathway would be to use a short diagnostic assessment before training, map results to NeedIndex bands, deliver targeted modules, and then reassess the specific concepts addressed. This would make professional development more accountable and iterative. Over time, aggregated item-level evidence could guide district-level or national training priorities, while still allowing schools to adapt activities to their local context and staff profile.

Table 7
Proposed tiered professional-development model based on NeedIndex findings

Tier	Main focus	Indicative content	Suggested activities	Evaluation
Tier 1	Foundational technical concepts	AI vs automation; narrow vs general AI; machine-learning workflow; AI techniques	Concept modules; misconception checks; matching tasks; visual ML pipeline activities	Objective pre/post items
Tier 2	AI capabilities, limitations, and uncertainty	Generalization limits; AI errors; uncertainty; atypical cases	AI-output evaluation; error analysis; classroom vignettes	Scenario-based interpretation tasks
Tier 3	Classroom-relevant pedagogical use	Feedback; assessment; differentiation; lesson planning; student support	Evaluation of AI-generated materials; redesign tasks; reflective discussion	Rubric-based judgement tasks
Tier 4	Ethics and governance	Privacy; bias; transparency; accountability; human oversight	Ethical dilemmas; school-policy cases; privacy checklists; classroom AI-use guidelines	Reflective/group task and implementation plan

7. Limitations and future directions

Several limitations should be considered. First, the study used a convenience sample recruited through open professional and personal networks; therefore, the findings should not be interpreted as fully representative of all Greek secondary education teachers. The sample may over-represent teachers with stronger interest in digital technologies, AI, or professional learning. Second, the cross-sectional design prevents causal inference. Although AI-specific training exposure was associated with higher objective AI literacy scores, the study cannot establish that training caused higher performance.

Third, the study measured objective conceptual AI knowledge, not pedagogical enactment. The test captures selected knowledge elements related to AI concepts, data, machine-learning processes, capabilities, limitations, and ethical issues, but it does not directly assess classroom practice, instructional design, situated ethical reasoning, or teachers' ability to integrate AI into teaching and learning. Fourth, although the instrument was translated, back-translated, and pilot checked for clarity, further measurement work is needed in the Greek context, including reliability across teacher subgroups, test-retest stability, sensitivity to professional-development interventions, and measurement functioning across relevant groups.

A further limitation concerns the breadth-depth trade-off of a 26-item objective test. Such an instrument is useful for efficient baseline mapping, but it cannot cover the full complexity of AI literacy, especially in relation to generative AI, multimodal systems, classroom policy, and student-facing ethical dilemmas.

Some concepts are represented by a small number of items, so item-level findings should be used as indicators of training need rather than definitive measurements of broad competence domains.

Future studies should combine objective knowledge testing with scenario-based and practice-based measures, such as classroom vignettes, AI-output evaluation tasks, ethical dilemma scenarios, interviews, lesson-design tasks, or observations. Intervention studies should also test whether NeedIndex-guided professional development reduces item-level gaps and whether conceptual gains are accompanied by more informed, responsible, and pedagogically meaningful AI-related practice.

8. Conclusions

Using an objective knowledge test, this study mapped AI literacy among Greek secondary education teachers and translated item-level performance into professional-development priorities. The findings indicate a moderate overall baseline, but also show that AI literacy is conceptually uneven. The largest knowledge gaps concerned foundational technical AI concepts: automation versus AI, narrow versus general AI, the machine-learning workflow, AI techniques, and limits in perception and generalization.

This pattern matters because a moderate average score should not be interpreted as adequate conceptual coverage. Teachers may show relatively strong performance on general, data-related, or ethical AI statements while still lacking technical foundations that support informed interpretation of AI-enabled systems. These foundations underpin teachers' ability to evaluate AI claims, understand system limitations, and engage responsibly with AI tools in school contexts.

The results also suggest that AI literacy is more strongly associated with professional and training-related variables than with general demographic characteristics. NeedIndex-based prioritization provides a practical bridge from assessment to professional-development design, supporting a staged approach that begins with technical foundations, moves to classroom-relevant interpretation of capabilities and limitations, and then deepens ethical and governance routines. Future research should examine whether improvements in objective AI knowledge translate into stronger pedagogical judgement and responsible classroom enactment.

Declarations

Ethics approval and consent to participate:

The study received institutional ethics approval from the Research Ethics Committee of the University of Macedonia, Greece (EHDE; meeting no. 13/6-6-2025; protocol no. 36/19-05-2025). The research was conducted in accordance with the ethical standards and guidelines of the Research Ethics Committee of the University of Macedonia, relevant institutional and national guidelines and regulations for research involving human participants, and the principles of the Declaration of Helsinki as applicable to non-clinical survey research. Participation was voluntary and electronic informed consent was obtained from all participants before questionnaire access. No directly identifying information was collected or stored.

Consent to participate:

Electronic informed consent was obtained from all participants before questionnaire access.

Consent to publish:

Not applicable. The manuscript reports anonymous, aggregated survey data and does not contain any identifiable individual data, images, or case details requiring consent for publication.

Clinical trial number

Not applicable.

Open access funding

Open access funding is intended to be covered through the University of Macedonia / HEAL-Link open access support route, subject to eligibility and confirmation.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

Nikolaos Nerantzis: Conceptualization; Methodology; Investigation; Data curation; Formal analysis; Writing - original draft; Visualization; Writing - review & editing. Dr Anastasia Psalti: Conceptualization; Methodology; Supervision; Writing - review & editing. Dr Kostas Zafiropoulos: Methodology; Formal analysis; Validation; Writing - review & editing.

Acknowledgement

The authors thank the teachers who voluntarily participated in the study.

Data Availability

De-identified data, a data dictionary, and analysis materials are available from the corresponding author upon reasonable request. The data are not publicly available due to ethics/privacy constraints and residual re-identification risk.

References

1. du Boulay B, Mitrovic A, Yacef K, editors. Handbook of artificial intelligence in education. Edward Elgar Publishing; 2023. https://doi.org/10.4337/9781800375413.
2. Jiang F, Li W, Luo C, Yin J, Kaynak O. What is artificial intelligence? An overview. Discover Artif Intell. 2022;2:4. https://doi.org/10.1007/s44163-022-00022-8.
3. Lu Y, Yu S. An introduction to artificial intelligence in education: bridging human and machine. Springer; 2021. https://doi.org/10.1007/978-981-16-2770-5.
4. Holmes W, Porayska-Pomsta K, Holstein K, Sutherland E, Baker T, Buckingham Shum S, et al. Ethics of AI in education: towards a community-wide framework. Int J Artif Intell Educ. 2022;32:504–26. https://doi.org/10.1007/s40593-021-00239-1.
5. UNESCO. AI and education: guidance for policy-makers. UNESCO; 2021.
6. UNESCO. Recommendation on the ethics of artificial intelligence. UNESCO; 2021.
7. Vuorikari R, Kluzer S, Punie Y. DigComp 2.2: The digital competence framework for citizens - with new examples of knowledge, skills and attitudes. Publications Office of the European Union; 2022.
8. Long D, Magerko B. What is AI literacy? Competencies and design considerations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery; 2020. pp. 1–16. https://doi.org/10.1145/3313831.3376727
9. Pinski M, Benlian A. AI literacy - towards measuring human competency in artificial intelligence. In: Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS 2023). 2023. https://doi.org/10.24251/HICSS.2023.021
10. Bender SM. Awareness of artificial intelligence as an essential digital literacy: ChatGPT and GenAI in the classroom. Chang Engl. 2024;31(2):161–74. https://doi.org/10.1080/1358684X.2024.2309995.
11. Kreinsen H, Schulz A. Towards a taxonomy of digital literacy competences for pre-service teacher education: a systematic review [OSF Preprint]. 2023. https://doi.org/10.35542/osf.io/xguzk
12. Ohlsson S. The role and function of theories in AIED. In: du Boulay B, Mitrovic A, Yacef K, editors. Handbook of artificial intelligence in education. Edward Elgar Publishing; 2023. pp. 31–44. https://doi.org/10.4337/9781800375413.00012.
13. Ding L, Kim S, Allday RA. Development of an AI literacy assessment for non-technical individuals: what do teachers know? Contemp Educational Technol. 2024;16(3):ep512. https://doi.org/10.30935/cedtech/14619.
14. Lintner T. A systematic review of AI literacy scales. npj Sci Learn. 2024;9:50. https://doi.org/10.1038/s41539-024-00264-4.

15. Sperling K, Stenberg C-J, McGrath C, Åkerfeldt A, Heintz F, Stenliden L. In search of artificial intelligence (AI) literacy in teacher education: a scoping review. Computers Educ Open. 2024;6:100169. https://doi.org/10.1016/j.caeo.2024.100169.
16. Crocker L, Algina J. Introduction to classical and modern test theory. Cengage Learning; 2008.
17. Ng DTK, Leung JKL, Chu SKW, Qiao MS. AI literacy: definition, teaching, evaluation and ethical issues. Proceedings of the Association for Information Science and Technology. 2021;58(1):504–509. https://doi.org/10.1002/pra2.487
18. Ng DTK, Wu W, Leung JKL, Chu SKW, Qiao MS. Design and validation of the AI literacy questionnaire: the affective, behavioural, cognitive and ethical approach. Br J Edu Technol. 2024;55(3):1082–104. https://doi.org/10.1111/bjet.13411.
19. Mishra P, Koehler MJ. Technological pedagogical content knowledge: a framework for teacher knowledge. Teachers Coll Record. 2006;108(6):1017–54. https://doi.org/10.1111/j.1467-9620.2006.00684.x.
20. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract. 2011;17(2):268–74. https://doi.org/10.1111/j.1365-2753.2010.01434.x.
21. Prilop CN, Mah D-K, Jacobsen LJ, Hansen RR, Weber KE, Hoya F. Generative AI in teacher education: educators' perceptions of transformative potentials and the triadic nature of AI literacy explored through AI-enhanced methods. Computers Education: Artif Intell. 2025;9:100471. https://doi.org/10.1016/j.caeai.2025.100471.
22. Asterhan CSC, Lefstein A. The search for evidence-based features of effective teacher professional development: a critical analysis of the literature. Prof Dev Educ. 2024;50(1):11–23. https://doi.org/10.1080/19415257.2023.2283437.
23. Darling-Hammond L, Hyler ME, Gardner M. Effective teacher professional development. Learn Policy Inst; 2017. https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report
24. Desimone LM. Improving impact studies of teachers' professional development: toward better conceptualizations and measures. Educational Researcher. 2009;38(3):181–99. https://doi.org/10.3102/0013189X08331140.
25. Morina F, Fütterer T, Hübner N, Zitzmann S, Fischer C. Effects of online teacher professional development on teacher-, classroom-, and student-level outcomes: a meta-analysis. Comput Educ. 2025;224:105247. https://doi.org/10.1016/j.compedu.2025.105247.

Figures

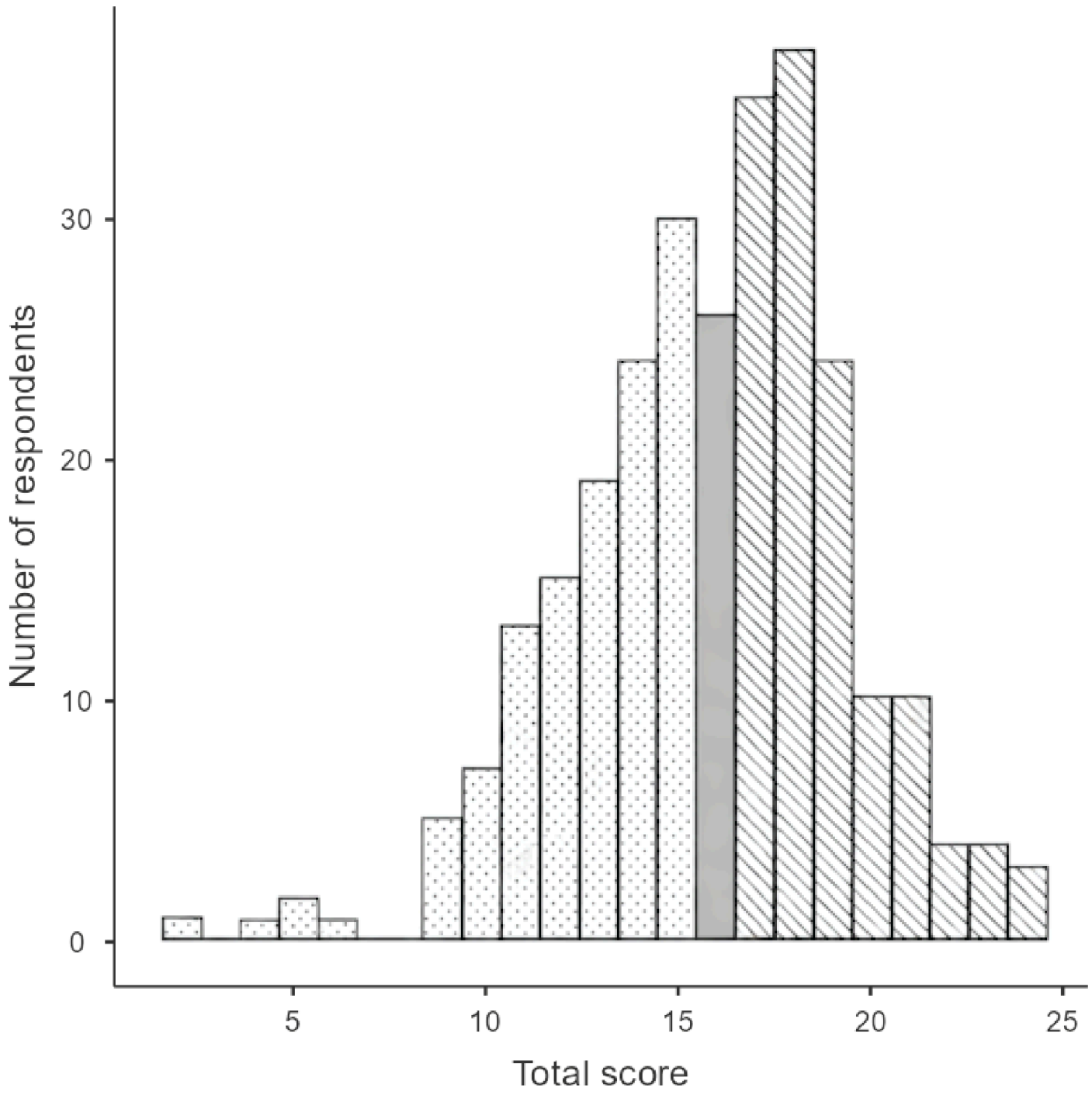


Figure 1

Distribution of total AI literacy test scores (0-26).

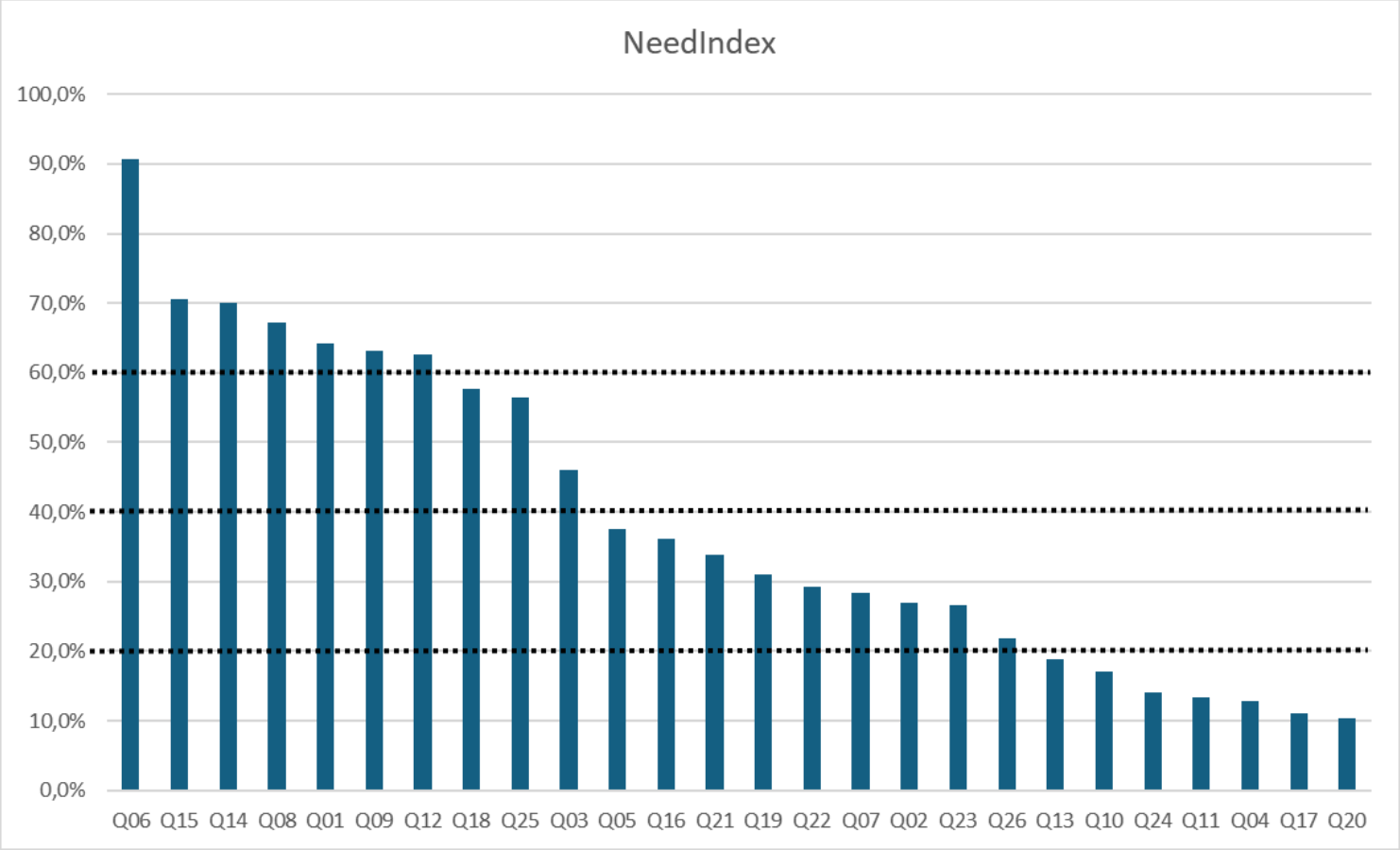


Figure 2

Item-level NeedIndex values ordered by training priority.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DEAdditionalanalysesareprovidedinESM1.docx