Journal of Turkish Science Education, 2026, 23(3), 554-579
DOI no: 10.36681/tused.2026.027

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

Digital transformation in zoology education: A methodology for training future biology teachers

Zukhra Bagimbayeva[1], Assiya Maimatayeva[2], Zhadyra Kabatayeva[3], Kuralay Alipina[4], Zhuldyz Tergenbayeva[5], Ruslan Safarov[6]

[1]*Sarsen Amanzholov East Kazakhstan University, Kazakhstan, ORCID ID: 0009-0002-5707-6973*
[2]*Abai Kazakh National Pedagogical University, Kazakhstan, ORCID ID: 0000-0002-4256-0802*
[3]*Sarsen Amanzholov East Kazakhstan University, Kazakhstan, ORCID ID: 0000-0003-2684-0086*
[4]*Sarsen Amanzholov East Kazakhstan University, Kazakhstan, ORCID ID: 0000-0002-5397-3215*
[5]*Sarsen Amanzholov East Kazakhstan University, Kazakhstan, ORCID ID: 0000-0001-8891-4066*
[6]*Faculty of Natural Sciences, L.N. Gumilyov Eurasian National University, Kh. Dosmukhamedov Atyrau University, Kazakhstan, Corresponding Author, safarov_rz@enu.kz, ORCID ID: 0000-0003-2158-6330*

ABSTRACT
The digital transformation of higher education has outpaced the methodological readiness of biology teacher training, where generic ICT courses rarely translate into discipline-specific teaching skills. This study develops and evaluates a 16-week "Digital Didactics of Zoology" module. It examined whether embedding digital tools within zoology pedagogy is associated with better outcomes than conventional instruction paired with separate ICT training. A mixed-methods quasi-experiment followed 84 pre-service biology teachers in two intact academic groups, an experimental group receiving the modüle (n = 42) and a control group receiving conventional teaching with a generic ICT course (n = 42). Data were derived from pre- and post-tests of subject knowledge, a 35-item competence questionnaire, case-based lesson design and micro-teaching tasks, blinded expert ratings, and coded lesson plans and interaction logs. Because (ANCOVA), allocation was not random, post-test scores were compared using analysis of covariance (ANCOVA), adjusting for the pre-test. The experimental condition scored **significantly** higher in zoology knowledge (adjusted difference 9.22 points, 95% CI [7.10, 11.33], d = 1.89) and in digital didactic competence (20.53 points, 95% CI [17.89, 23.17], d = 3.38) and higher on lesson quality (g = 1.54) and expert evaluations (g = 2.48). An exploratory structural equation model (SEM) fitted to the full sample (CFI = .994, RMSEA = .060, SRMR = .027) associated digital proficiency with pedagogical reasoning (β = .71, bootstrap 95% CI [.55, .83]); this association appeared in both conditions, and the two estimates were not compared formally. The module provides teacher-education programs with a transferable design for technology-integrated zoology instruction, pending replication across institutions.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
09.01.2026
Accepted:
21.09.2026
Available Online:
26.09.2026

KEYWORDS:
Biology teacher education, TPACK, digital didactics of zoology, virtual dissection, Geographical Information Systems (GIS).

To cite this article: Bagimbayeva, Z., Maimatayeva, A., Kabatayeva, Z., Alipina, K., Tergenbayeva, Z., & Safarov, R. (2026). Digital transformation in zoology education: a methodology for training future biology teachers. *Journal of Turkish Science Education*, *23*(3), 554-579. http://doi.org/10.36681/tused.2026.027

Introduction

Digital transformation has reshaped teaching across higher education, yet biology sits awkwardly within this shift. The subject matter of biology is visual and dynamic, spanning scales from the cell to the ecosystem (Neumann & Waight, 2020; Tanner & Allen, 2006). In particular, zoology requires students to build spatial models of anatomy (Sharmin et al., 2023), weigh the ethical implications of dissection (Ormandy et al., 2022), and identify and monitor organisms in the field. Generic educational technology does not, on its own, meet these discipline-specific demands.

Teacher-preparation programs tend to leave a technological-pedagogical disconnect (de Oliveira et al., 2019; Jensen, 2011). Pre-service teachers acquire general digital literacy (presentation software, learning-management systems) yet often lack the Technological Pedagogical Content Knowledge (TPACK) to fold discipline-specific tools such as virtual laboratories, GIS, or AI identification apps into inquiry-based lessons (Irdalisa et al., 2022; Jiménez Sierra et al., 2023; Valtonen et al., 2020). The disconnect widens where ICT training runs separately from disciplinary content (Hajiyeva, 2021), a pattern common in rapidly digitalizing systems in Central Asia, including Kazakhstan (Abiltayeva et al., 2025; Smagulova & Kalieva, 2025), as well as in other developing contexts (Chisingui & Costa, 2020; Diem et al., 2024). Where technology is added as a supplement rather than used as a way of knowing, adoption stays fragmented, and student outcomes change little.

This study adopts a distinct approach. Rather than treating digital skills as an isolated competence, it embeds them directly within the pedagogy of zoology, ensuring that each tool is acquired through the specific educational challenge it addresses: AI-driven identification to address the constraint that routine species determination places on field and classroom work (Gaston & O'Neill, 2004; Wäldchen & Mäder, 2018), virtual dissection to navigate the ethical and practical constraints of specimen work (Telecan et al., 2025), and GIS alongside 3D modeling to overcome the difficulties of representing space and scale. The novelty lies not in the familiarity of the individual tools but in their integration into a single evaluated module, which is rigorously tested against an active control group that receives identical zoological content coupled with a separate, conventional ICT course. This design effectively isolates the impact of situating technology within a specific discipline – highlighting the distinction between merely knowing a tool and understanding how to use it to instruct.

The primary objective of this research is to develop a discipline-specific "Digital Didactics of Zoology" module and experimentally validate it against traditional zoology instruction supplemented by generic ICT training. In this context, validation entails a controlled comparison: the module is evaluated not by student engagement with digital tools, but by whether embedding these technologies within zoological pedagogy yields measurable learning outcomes that standalone ICT courses fail to produce. Four research questions guide the analysis:

RQ1. Does the module produce larger gains in zoology subject knowledge than conventional instruction with separate ICT training?

RQ2. Does it raise digital didactic competence, understood as the capacity to select, adapt, and troubleshoot digital tools for specific zoology topics?

RQ3. Does it improve pedagogical readiness and the quality of technology-integrated lesson design?

RQ4. Within the experimental condition, are pre-service teachers' digital-tool proficiencies directionally linked to their pedagogical reasoning?

The hypotheses are directional. H1–H3 predict higher post-intervention performance in the experimental group for subject knowledge, digital didactic competence, and pedagogical quality, respectively. H4 predicts that proficiency with specific tools (3D manipulation, GIS analysis, and virtual experimentation) is positively associated with domain-matched pedagogical reasoning, rather than standing apart as an isolated technical skill.

Literature Review

TPACK in Biology Education

Digital tools, such as virtual laboratories and 3D models, enhance engagement and retention in zoology (Sautière et al., 2019). However, their integration into teacher preparation remains uneven, particularly in disciplines where pedagogy must align with subject-specific affordances (Ambarwati et al., 2019). The TPACK framework was developed to close this gap by treating effective technology use not as a standalone skill, but as the intersection of content, pedagogical, and technological knowledge. Under this framework, a biology teacher must understand how a particular tool represents phenomena that are otherwise abstract or invisible (Jiménez Sierra et al., 2023). Teachers with well-developed TPACK integrate inquiry and higher-order thinking more effectively than peers who are digitally literate but pedagogically untrained (Valtonen et al., 2020). The shortfall is specific rather than general. Among 102 science teachers, the design dimension of TPACK was the weakest, use stayed at the level of smart boards and presentation software, and insufficient infrastructure and inadequate training were named as the obstacles to deeper integration (Dinçer, 2024).

Furthermore, recent research on interdisciplinary zoology instruction notes that teachers now need to connect biological, computational, and environmental reasoning simultaneously (Anapana et al., 2025). This logic is discipline-bound; for instance, the reasoning required to teach habitat mapping with GIS differs significantly from that needed to run a cell-division simulation (Neumann & Waight, 2020). Consequently, teacher preparation programs that treat digital competence as supplementary rather than foundational leave this specialized reasoning undeveloped (Hsu et al., 2023; Pererva et al., 2020). This broader trend is documented in a systematic review of ten years of mobile learning in biology education. By screening 1,315 records and analyzing 66 records, Sangur et al. (2025) reported steady growth in mobile tool usage, primarily for visualizing biological concepts and supporting practical work.

Virtual Laboratories and Ethical Alternatives

Virtual dissection is adopted for two reasons: one ethical and one methodological. Ethically, it removes the need to euthanize animals for single-use study; a meta-analysis in The American Biology Teacher found that non-animal methods matched or outperformed traditional dissection in most comparisons and eased the discomfort some students feel with real specimens (Ormandy et al., 2022), which supports the 3R principle (replacement, reduction, and refinement). Methodologically, a virtual specimen can be cut, undone, and recut without limit, whereas a physical specimen can be used only once and cannot be restored (Kavvadia et al., 2023). Because a mistaken incision no longer destroys the material, students practice to fluency, and their ability to isolate and rotate organ systems in three dimensions improves their performance on anatomical tests over wet specimens (Telecan et al., 2025). A randomized trial in medical anatomy teaching points in the same direction: students who worked through a virtual dissection before the gross dissection performed better on the subsequent anatomy test than those who prepared from textbooks, although the difference fell short of conventional significance (Boscolo-Berto et al., 2021).

Visualization Technologies: 3D, AR, and VR

Traditional zoology teaching represents three-dimensional structures through flat diagrams, forcing students to reconstruct spatial relationships in their working memory. Learners with weaker spatial ability struggle the most, and their early misconceptions tend to persist (Anniesmitha et al., 2025; Sharmin et al., 2023). Interactive 3D, AR, and VR technologies shift this cognitive reconstruction into the interface; features such as rotation, isolation, and cross-sectioning supply the depth cues that static atlases lack (Reinke et al., 2021). Furthermore, manipulating objects in virtual space engages spatial perception processes more effectively than their 2D equivalents (Tung & Chang, 2024). Immersive environments also lower extraneous cognitive load by allowing learners to explore anatomy from the inside out (Kim & Choi, 2025; Majewska & Vereen, 2023). Meta-analytic evidence for AR in biology demonstrates higher engagement and conceptual understanding, with the largest gains observed in students who begin with weak spatial reasoning. Consequently, this technology narrows rather than widens the performance gap (Sattar et al., 2025). Even so, its adoption in teacher-preparation curricula remains in the early stages, and few programs integrate these tools with subject-specific pedagogy within a single framework (Biletska et al., 2021).

Geospatial and AI-Based Tools in Field Zoology

Two of the four tool families in the module address learning outside the laboratory, where the object of study is a distribution rather than a structure. Direct evidence on geospatial tools in biology teacher preparation remains scarce, and the relevant findings come largely from geography education at the school level, so they are read here as an adjacent rather than a direct evidence base. Web GIS makes demographic and distributional data workable in school inquiry, and it carries an evaluative requirement with it: because these platforms accept contributions from anyone, the layers on offer range from undocumented to authoritative, so provenance has to be checked before a map is used for teaching (Kerski, 2019). Reviewing 36 empirical studies of inquiry teaching with geospatial technologies, Mašterová (2023) reports that students develop both digital competence and knowledge of the topic addressed through the tool, but that prior teacher preparation is a condition for this outcome rather than an accompaniment to it. There is also evidence that geospatial reasoning is a construct in its own right and not a by-product of exposure to maps: in a sample of 1,037 upper-secondary students, geospatial thinking partially mediated the relationship between family capital and achievement in geography, accounting for 24.68% of the total effect (Zhang et al., 2023).

Species identification is itself an object of teaching, and the skills it calls for are treated in biology education as a prerequisite for understanding biodiversity and ecology, with the choice of specimens, keys, and preparatory classroom work shaping what pupils take from a field task (Randler, 2008). Competence in teaching this material is itself built rather than assumed: across 151 pre-service science teachers, both knowledge of biodiversity and perceived competence to teach it rose with the amount of content and pedagogical coursework completed (Tergenbayeva et al., 2023). The identification tools raise a parallel question about evaluating output. Platforms such as iNaturalist pair automated suggestion with community verification, and that verification is partial in practice; of some 445,000 observations recorded in one state-level review, about 40% had reached verified Research Grade status, and less conspicuous organisms were underrepresented (Baker et al., 2023). A PRISMA review of iNaturalist in biology education finds uses clustered around BioBlitz activities, outdoor education, and herbarium work, and closes with guidelines for preparing the application for school use (Rode & Torkar, 2023). Drawing on interviews with 31 participants aged 7 to 20, Herodotou et al. (2024) identify the forms of online and field participation that support science learning rather than data collection alone, and set out a blended framework for biodiversity monitoring on that basis. Reviews of artificial intelligence in biology education record steady growth in reported applications alongside uneven topic coverage and few studies of AI-supported games and simulations (Duran & Dikmenli, 2025). Both lines converge on the competence this module treats as the object of instruction rather than its precondition, since an automated identification is a provisional taxonomic claim that a teacher must be able to test against diagnostic characters before it enters a lesson.

Contextual Factors in Teacher Education

What "digital competence" means is also shaped by policy and setting. General practice tends to overlook the epistemic particularities of scientific disciplines, thereby limiting its practical effect (Garzón-Artacho et al., 2021; Nietbaeva, 2023). Biology teacher education in Central Asia faces a dual task: raising overall digital literacy while preparing teachers for subject-specific tools. Work in Kazakhstan reports a strong policy commitment to digital transformation, alongside methods courses in which digital tools remain peripheral or are used only for presentation purposes (Abiltayeva et al., 2025; Smagulova & Kalieva, 2025). Similarly, research in Angola links technology-enhanced biology teaching to sustainability goals while noting the same absence of systematic, subject-focused training (Chisingui & Costa, 2020).

Assessing Digital Competence and Its Limits

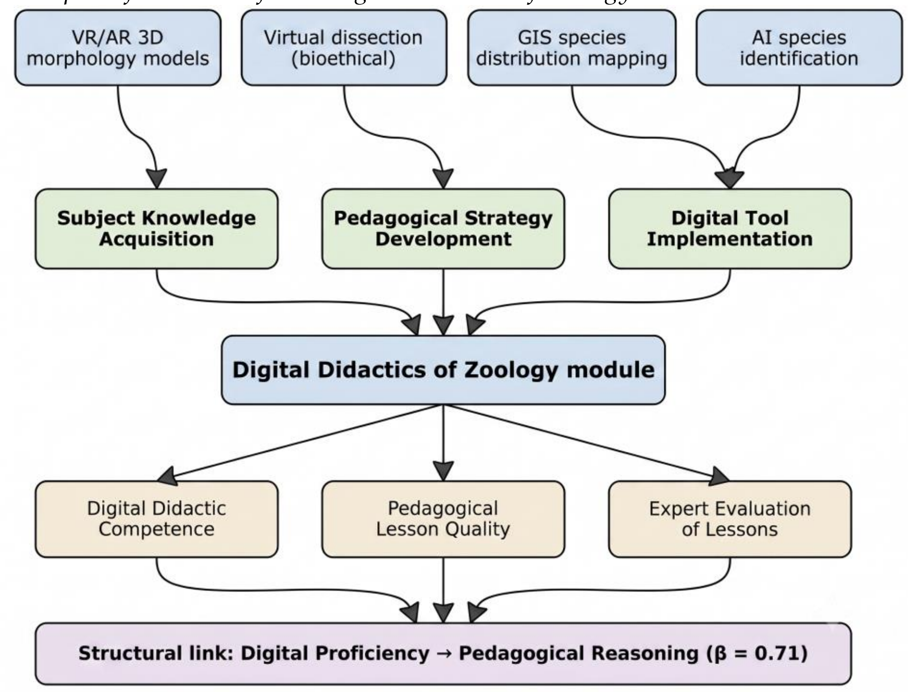
Teachers' digital competence has mostly been assessed with TPACK and DigCompEdu, the latter specifying 22 competences across six areas (Ghomi & Redecker, 2019), and these frameworks underlie many self-report surveys (Tzafilkou et al., 2023). Adaptations for biology capture self-reported data literacy, content creation, and tool use; the Kazakhstani study cited above, for example, found moderate technological knowledge and positive attitudes, but little experience with virtual laboratories or interactive animal biology resources (Abiltayeva et al., 2025). The problem is that these instruments are domain-agnostic. They distinguish broad categories of competence but not, say, using a presentation tool from facilitating inquiry with a virtual dissection platform. Furthermore, they lean on access, frequency, and self-efficacy rather than on discipline-specific practice, such as modeling ecological systems or interpreting structure–function relationships (Adhani & Nursia, 2024). A candidate can therefore score high on a general scale while being unable to orchestrate a technology-rich zoology lesson. Recent work instead calls for context-based competence, measured through tasks that reflect authentic zoology teaching and are tied to taxonomy, morphology, physiology, and behavior rather than to tools treated as interchangeable (Kang et al., 2024; Nurbaya, 2023; von Kotzebue, 2022). Validated zoology-specific instruments of this kind remain scarce. Standards documents describe the same territory in equally general terms, the ISTE standards for teachers being a long-established example (International Society for Technology in Education [ISTE], 2002). Reviews of digital pedagogy reach that conclusion from the research side: surveying the literature, Väätäjä and Ruokamo (2021) separate pedagogical orientation, pedagogical practice, and digital competence as three dimensions of digital pedagogy, none of which is specified for a particular subject.

Conceptual Framework

These strands converge on a single design problem: how to embed digital competence within specific disciplines rather than treating it as a generic skill. Figure 1 sets out the conceptual framework that guided the module. It is structured around three components that the literature traditionally treats as distinct, but which the module integrates: subject-knowledge acquisition, pedagogical-strategy development, and digital-tool implementation. Each digital tool is introduced through the specific pedagogical problem it addresses; for instance, AI identification is contextualized against the constraint of routine species determination, virtual dissection against the limitations of specimen work, and GIS and 3D models against the challenges of representing space and scale. These three components inform the module, whose impact is evaluated through three measured outcomes: digital didactic competence, pedagogical lesson quality, and expert-rated lesson design. The framework also posits the relationship tested in RQ4: that proficiency with discipline-specific tools and pedagogical reasoning covary, rather than developing as unrelated, parallel skills. This study operationalizes this framework and tests it against an active control group.

Figure 1

Conceptual framework of the "Digital Didactics of Zoology" module



Methods

Research Design and Participants

The study used a mixed-methods quasi-experimental design with pre- and post-testing. The quantitative and qualitative strands were conducted over the same period and given equal weight, and their results were integrated at the interpretation stage, making the design convergent. The qualitative material was collected alongside the outcome measures rather than after them, and its role was to describe how lesson design differed between the two conditions, not to generate hypotheses for later testing. Integration is located in three places: the coding frame in Table 3, which reports code frequencies separately by condition; the paragraph on archived lesson plans in the Results section, where rubric scores and coded artifacts are read against each other; and the Discussion section, where each research question is answered from both strands.

Fieldwork ran throughout the 2022–2023 academic year at a public pedagogical university during the fourth semester, when the core zoology course is taught. Participants were 84 preservice biology teachers in the second year of the teacher-education program. Eligibility required current enrolment in the zoology course and in the teacher-education track. No further exclusion criterion was applied. The program admits only full-time students, so no part-time or distance learners had to be excluded, and the two parallel groups taking the course that semester therefore constituted the entire eligible population rather than a sample drawn from it. All 84 students in those groups provided consent and completed both measurement points, so 84 cases were included in the analysis. There was no attrition, exclusion after allocation, or missing observations on any measure; therefore, no imputation was required. The term "preservice biology teachers" is used throughout for this group.

The assignment was made at the level of intact academic groups rather than at the level of individual students. The two groups that took the zoology course in parallel were assigned to the experimental and control conditions, with one group to each, yielding 42 participants per condition.

Allocation was therefore not random, which is why the design is described as quasi-experimental and why no randomization procedure is claimed. Because allocation was not random, comparability at entry was examined rather than assumed. Table 1 reports standardized mean differences with confidence intervals for age, grade-point average, baseline ICT skills, both pre-tests, and the baseline expert rating. The largest standardized difference was 0.12, and every interval included zero; however, intervals of this width with 42 cases per condition do not establish equivalence, so the primary analyses adjust for the pre-test rather than rely on baseline balance.

Both conditions were taught by the same instructor and received the same number of contact hours, one 90-minute session per week for 16 weeks, and the same assessments, which held instructor effects and time-on-task constant. The control group did not take the digital instruction. It received a generic ICT course of equivalent length in the same semester, covering office applications, work in the learning-management system, information search and processing, and the general methodology of e-learning, none of it tied to biological content. The contrast is therefore between discipline-embedded and generic digital preparation rather than between digital preparation and none.

Two design features guard the final performance task against favoring the experimental condition by construction. Both conditions completed the task in the same virtual laboratory environment and received the same introductory session on the environment beforehand, so familiarity with the assessment platform did not differ between conditions. The task itself asked for a diagnostic biological problem to be solved rather than for a particular tool to be operated, which keeps the outcome at the level of reasoning rather than of tool fluency. The experimental condition nevertheless had longer exposure to tools of this class over the term, and the Limitations section returns to what that implies.

No procedural measure was taken to keep the two conditions apart during the term, and in ordinary timetabled teaching, neither temporal nor spatial separation was available. What limited the exchange of materials was the absence of salience rather than any barrier: no group was described to its members as receiving special treatment, and no attention was drawn to particular elements of a lesson, so those elements were taken as an ordinary part of the course and gave students little reason to pass them to peers in the parallel group. Exchange was possible and cannot be ruled out, and the Limitations return to it. As part of informed consent, participants in both conditions were told that two versions of the course were being compared, without being told which version was expected to perform better. Implementation fidelity was monitored with a weekly checklist, on which the instructor recorded which planned elements of each session had actually been delivered; this self-report was cross-checked against non-participant observation of one session in four by a member of the research team who taught neither condition. Observation focused on the sessions in which the module prescribed a specific digital tool, since that is where any departure from the plan would matter most, and the visits were conducted in the format of the scheduled peer classroom visits that the department carries out in any case. Keeping them in that familiar format was deliberate: singling those sessions out as part of a study would have made students self-conscious and changed the very teaching the observation was meant to record. Adherence exceeded 90% across the term.

The study was conducted within the regular curriculum of the zoology course. It involved no procedure beyond the teaching and assessment methods ordinarily used in the program, no biomedical or clinical intervention, and no collection of sensitive personal data, and it was not submitted to a separate research-ethics committee. Participation was voluntary, and all participants gave written informed consent after being told the purpose of the study, the procedures involved, and their right to decline or to withdraw at any point; no course grade or academic standing depended on participation, and students who had not consented would have followed the same course without their data being used. Responses were pseudonymised at collection, stored on institutional servers with access restricted to the research team, and are reported only in aggregate, so that no individual participant can be identified from anything published here.

Table 1

Participant characteristics and comparability of the two conditions at entry (N = 84)

Characteristic	Experimental (n = 42)	Control (n = 42)	g [95% CI]
Male, n (%)	7 (16.7)	6 (14.3)	RD 0.02 [-0.13, 0.18]
Female, n (%)	35 (83.3)	36 (85.7)	
Age, years	19.40 (0.90)	19.50 (0.80)	-0.12 [-0.55, 0.31]
Grade-point average (max 4.0)	3.42 (0.31)	3.38 (0.35)	0.12 [-0.31, 0.55]
Baseline ICT score (0–100)	62.50 (8.40)	61.80 (9.10)	0.08 [-0.35, 0.51]
Zoology knowledge pre-test (0–100)	44.80 (7.09)	45.30 (6.80)	-0.07 [-0.50, 0.36]
Digital didactic competence pre-test (0–100)	41.50 (8.20)	42.10 (7.90)	-0.07 [-0.50, 0.35]
Baseline expert rating (1–5)	2.90 (0.60)	2.90 (0.60)	-0.00 [-0.43, 0.43]

Note. Values are M (SD) unless marked otherwise. The final column gives Hedges' g for the standardized difference between conditions with its 95% confidence interval; a positive value favours the experimental condition. For sex the entry is the difference in the proportion of men, $\chi^2(1) = 0.09$, $p = .763$, Cramér's V = .03. Intervals that include zero do not establish equivalence at entry, which is why the primary analyses adjust for the pre-test.

Intervention Structure

The experimental group worked through a 16-week module built on the three components of the conceptual framework (Figure 1): subject-knowledge acquisition, pedagogical-strategy development, and digital-tool implementation. Each weekly session paired a short theoretical input with a hands-on workshop using one tool family: (1) virtual dissection laboratories with haptic feedback; (2) VR/AR-enabled 3D animal-morphology models; (3) GIS-based species-distribution mapping; and (4) AI-driven species-identification applications.

A week-by-week comparison of the experimental and control syllabi is given in Appendix A, and a sample worksheet for each tool family is provided in Appendix B so that the practical form of the intervention can be inspected directly. The control group covered the same zoology content through lectures, textbook study, and physical-specimen work, with the generic ICT sessions kept separate from the biology curriculum.

Assessment Framework and Instruments

Competence was assessed in three domains, each with a dedicated instrument. Descriptive statistics (pre- and post-intervention means and standard deviations) for every domain are reported in Table 2 before the inferential analysis.

Subject knowledge was measured with a 40-item test that combined multiple-choice questions on taxonomy, anatomy, and ecology with four open-ended items requiring students to translate a biological concept into a teaching task (for example, outlining how they would explain insect metamorphosis to a mixed-ability class). Three zoology educators rated every item for relevance to establish content validity; items with a content-validity index below 0.80 were revised or removed. Internal consistency was acceptable (Cronbach's $\alpha = .82$). Two raters scored the open-ended items independently against an analytic rubric (Appendix C), with high agreement (ICC = .89).

Digital didactic competence was measured with a 35-item questionnaire ($\alpha = .87$) spanning three subscales: appropriateness of tool selection for a given zoology topic (5-point Likert items), depth of pedagogical rationale in short written responses, and technical troubleshooting in scenario items. Sample items and the scoring key appear in Appendix C.

Pedagogical readiness was assessed through case-based tasks. Students responded to authentic classroom scenarios with a lesson plan, a 10-minute micro-teaching episode, and a peer-review exercise. Lesson plans and micro-teaching were scored with analytic rubrics covering conceptual accuracy, pedagogical appropriateness, technological integration, and accessibility (Appendix C). Two trained raters worked from the same rubric. Before scoring began, they went through it together on several sample pieces of work, comparing their judgements and settling differences of interpretation, and only then scored independently. Approximately 20% of the material was scored by both raters, that is, 34 of the 168 lesson plans and 17 of the 84 micro-teaching episodes, and the remainder was scored by one rater. Raters worked from anonymised material and did not know which condition a plan or episode came from, who had written it, or whether it belonged to the pre- or the post-intervention set, and they were not told the hypotheses of the study. Blinding to condition was imperfect in one respect that should be stated plainly: a lesson plan from the experimental condition often names a specific digital tool, so a rater could sometimes infer the condition from the content of the plan even with every label removed. Agreement on the doubly scored material was satisfactory. For the open-ended subject-knowledge items the intraclass correlation was .89, from a two-way mixed-effects model for absolute agreement with single measures; the two raters being the only raters of interest rather than a sample of raters. The ordinal rubrics gave κ = .84 for lesson plans and κ = .81 for micro-teaching; these are unweighted coefficients, and the limitations note that a weighted coefficient would suit an ordinal rubric better.

Data Collection Procedures

Baseline data were collected in week 1, and outcome data in the week after the module ended. Before the main study, all instruments were piloted with a separate group of 20 students to assess clarity and timing, and minor wording changes followed. Throughout the term, students' instructional artifacts (lesson plans, virtual-reality models, reflective notes) were archived, together with interaction logs from the virtual-laboratory sessions, for the qualitative strand.

Qualitative Analysis

The qualitative strand examined why the two conditions designed lessons differently. Its sources were the written comments of the three expert evaluators, the 168 lesson plans, and the virtual-laboratory interaction logs. The 168 plans are one plan per participant at each of the two measurement points, 84 before and 84 after the module, so the two conditions contribute equally to the corpus, and percentages reported for lesson features use the 42 plans of the relevant condition and measurement point as their denominator. Interaction logs are available for every session the experimental condition attended and for the final performance task, which both conditions completed, and only the logs of that shared task are used for comparison between conditions. The quotations attributed to experts come from the evaluators' written comments on individual lessons. Micro-teaching episodes were rated with the same rubric as the plans but were analysed separately, and their ratings entered the pedagogical-readiness measure rather than the plan corpus. The unit of analysis was the idea-bearing segment. Two researchers coded a common set of 40 plans to build a codebook, then coded the full corpus independently, resolving disagreements by discussion. The codebook was fixed before the full corpus was coded, and three of the codes record the absence of a practice rather than its presence so that the frame could register evidence running against the expected direction. Both coders are authors of this study and took part in designing the module. That gave them the disciplinary and pedagogical knowledge the coding required and, at the same time, an interest in what the coding would show, and the safeguards described here follow from the second of those facts as much as from the first: the counter-codes were introduced deliberately, the codebook was closed before the corpus was opened, coding was carried out on anonymised material, and the frequency of every code in each condition is published in Table 3, so that a reader can weigh the frame itself rather than take the interpretation on trust. Coding yielded 14 descriptive codes across 312 segments, which were grouped into four themes: contextualisation of technology, inquiry orientation, accessibility, and conceptual precision. Intercoder reliability was acceptable (Krippendorff's α = .82). Frequencies from the rubric-based expert ratings and the interaction logs are reported quantitatively in the Results; the thematic analysis interprets those patterns and is illustrated with representative excerpts. The full coding frame, with the frequency of every code in the two groups, is reported in Table 3.

Analytical Approach

Quantitative analysis was carried out in IBM SPSS Statistics 26, with the structural model estimated in the semopy package for Python and bootstrap intervals obtained in the same environment. Four outcomes were designated primary: zoology subject knowledge, digital didactic competence, pedagogical lesson quality, and the blinded expert evaluation. Every other comparison reported below, including the subscale contrasts and the structural model, is exploratory and is labelled as such where it appears. Familywise error across the four primary contrasts is controlled with the Holm procedure. Confidence intervals are 95% throughout unless stated otherwise, and bootstrap intervals use 2,000 resamples.

For the two outcomes with a pre-test on the same scale, the primary between-group model is an analysis of covariance with the post-test score as the dependent variable, the pre-test score as the covariate, and condition as the fixed factor. This model replaces the comparison of gain scores used in the previous version. The reason is the allocation procedure: with intact groups rather than randomisation, adjustment for the pre-test is required rather than optional. Homogeneity of regression slopes was tested by adding the interaction of condition and covariate, and the residuals were examined with the Shapiro–Wilk test and Levene's test; the outcome of each check is reported with the model it belongs to. Adjusted means with their intervals, the adjusted difference with its interval, partial eta squared, and the standardised adjusted difference are reported for every model. Gain-score comparisons are retained as sensitivity analyses.

Pedagogical lesson quality was measured only after the intervention, so no covariate is available for it, and it is compared with Welch's t-test; for the same reason, no change over time is claimed for that outcome, only a difference between conditions at the end of the module. The expert evaluation is also reported with Welch's t-test, because its residual variances differed by condition and its baseline rating carried no information about the outcome, which makes the covariance model inappropriate for it; the analysis of covariance was run for that outcome as a check and led to the same conclusion. Both are accompanied by the mean difference and Hedges' g, each with an interval. Within-condition change from pre-test to post-test is reported with paired-samples t-tests, the mean change with its interval, and the standardised mean change. Where a parametric assumption failed, the Mann–Whitney U test is reported alongside the parametric result, with the convention for U stated explicitly and with rank-biserial correlation and Cliff's delta as effect sizes.

Research question 4 was addressed with a structural equation model, which is reported as exploratory. Two latent variables were specified: Digital Proficiency, indicated by 3D-manipulation, GIS-analysis, and virtual-experimentation scores, and Pedagogical Reasoning, indicated by spatial-reasoning, ecosystem-teaching, and inquiry-lesson scores. A single structural path linked the first construct to the second. The model is deliberately parsimonious, with six indicators, two factors, and thirteen free parameters, and estimation used maximum likelihood on the full sample. Because 84 cases are modest for covariance-based modelling, fit is reported with the full index set recommended for small samples rather than with a single statistic, and the structural coefficient is accompanied by a percentile bootstrap interval. The coefficient was also estimated separately within each condition, but the two estimates are reported side by side as descriptive information only: a formal comparison would require a multi-group model with measurement invariance established across conditions, which 42 cases per condition will not support, and no such comparison is made. Because tool proficiency and pedagogical reasoning were measured at the same point in time, the model is read as a set of associations, and no temporal or causal ordering is inferred from it.

The qualitative and quantitative strands were then read against each other so that the test and survey results, the rubric scores, and the coded artifacts could corroborate or qualify one another.

Results

Results are organised by research question. Table 1 reports comparability at entry, and Table 2 reports descriptive statistics for every domain together with the between-group comparisons, so the data can be read before the inferential detail. The experimental condition scored higher on all four questions, although the size of the difference varied. Because allocation was not random and the measures within a wave were taken at the same time, the findings are stated as differences between conditions and as associations, and no causal or temporal ordering is claimed.

Table 2

Outcomes by condition, with within-condition change and between-condition comparisons

Outcome and condition	Pre M (SD)	Post M (SD)	Change M [95% CI]	Adjusted post M [95% CI]	Between-condition comparison
Subject knowledge (0–100), Experimental	44.80 (7.09)	73.20 (7.57)	28.40 [26.78, 30.02]	73.40 [71.91, 74.90]	$F(1, 81) = 75.02$, $p < .001$, $d = 1.89$, $\eta^2 p = .481$
Control	45.30 (6.80)	64.39 (7.36)	19.09 [17.60, 20.59]	64.19 [62.69, 65.68]	
Digital didactic competence (0–100), Experimental	41.50 (8.20)	76.71 (11.18)	35.21 [33.09, 37.33]	77.02 [75.16, 78.89]	$F(1, 81) = 239.55$, $p < .001$, $d = 3.38$, $\eta^2 p = .747$
Control	42.10 (7.90)	56.81 (9.64)	14.71 [13.09, 16.33]	56.49 [54.63, 58.36]	
Pedagogical lesson quality (1–5), Experimental	–	4.18 (0.57)	–	–	$t(74.2) = 7.14$, $p < .001$, $g = 1.54$ [1.06, 2.03]
Control	–	3.10 (0.80)	–	–	
Expert evaluation (1–5), Experimental	2.90 (0.60)	4.46 (0.44)	1.56 [1.31, 1.81]	–	$t(68.9) = 11.46$, $p < .001$, $g = 2.48$ [1.91, 3.05]
Control	2.90 (0.60)	3.00 (0.70)	0.10 [-0.20, 0.40]	–	

Note. Values are M (SD) unless marked otherwise; – marks a quantity that does not exist for that outcome. Change is the paired difference from pre-test to post-test. Adjusted post-test means come from an analysis of covariance with the pre-test as covariate and condition as fixed factor, and the effect size in those rows is the standardized adjusted difference with partial eta squared. Pedagogical lesson quality has no pre-test, and the expert evaluation had unequal residual variances with an uninformative covariate, so both are compared with Welch's t-test and reported with Hedges'g. The digital-competence contrast was confirmed without distributional assumptions on the gain scores, $U = 1742$ for the experimental relative to the control condition, $p < .001$, rank-biserial $r = .97$. All four primary contrasts remain significant after Holm correction.

RQ1 Subject Knowledge

Scores rose in both conditions between the pre-test and the post-test (experimental M change = 28.40, 95% CI [26.78, 30.02], $t(41) = 35.38$, $p < .001$, $d_z = 5.46$; control M change = 19.09, 95% CI [17.60, 20.59], $t(41) = 25.77$, $p < .001$, $d_z = 3.98$). With the pre-test entered as a covariate, the adjusted post-test mean was 73.40, 95% CI [71.91, 74.90], in the experimental condition against 64.19, 95% CI [62.69, 65.68], in the control condition, an adjusted difference of 9.22 points, 95% CI [7.10, 11.33], $F(1, 81) = 75.02$, $p < .001$, partial $\eta^2 = .481$, $d = 1.89$. The model assumptions held: the regression slopes did not differ by condition ($p = .821$), and the residuals were normally distributed ($p = .762$) with homogeneous variances ($p = .816$). The gain-score comparison retained as a sensitivity analysis leads to the same conclusion (difference = 9.30, 95% CI [7.13, 11.47], $g = 1.84$). The difference was widest on application items that asked students to combine biological content with a teaching strategy (experimental M change = 32.7 against control M change = 18.9). Higher accuracy on anatomical-reasoning items (87% against 62%) and on ecological-system modelling (79% against 54%) locates the difference in spatial and systems reasoning rather than in recall alone.

RQ2 Digital Didactic Competence

Scores rose in both conditions (experimental M change = 35.21, 95% CI [33.09, 37.33], t(41) = 33.54, p < .001, d_z = 5.18; control M change = 14.71, 95% CI [13.09, 16.33], t(41) = 18.34, p < .001, d_z = 2.83). Adjusted for the pre-test, the post-test mean was 77.02, 95% CI [75.16, 78.89], against 56.49, 95% CI [54.63, 58.36], an adjusted difference of 20.53 points, 95% CI [17.89, 23.17], F(1, 81) = 239.55, p < .001, partial η^2 = .747, d = 3.38. The regression slopes were homogeneous (p = .737), and the residual variances did not differ (p = .055), but the residuals departed from normality (p < .001), so the contrast was repeated without distributional assumptions. On the gain scores, the Mann–Whitney test gives U = 1742 for the experimental relative to the control condition, the complementary statistic being 22 of a possible 1,764, p < .001, with rank-biserial r = .97 and Cliff's delta = .97, 95% CI [0.91, 1.00]. The gain-score t-test agrees (difference = 20.50, 95% CI [17.87, 23.13], g = 3.36). All three subscales moved in the same direction, with tool selection changing by 38.1 points against 12.3, depth of pedagogical rationale by 33.8 against 11.9, and technical troubleshooting by 31.5 against 9.4; these subscale contrasts are exploratory and are not corrected for multiplicity. The difference was clearest where a scenario required adapting a resource for diverse learners (82% against 43% optimal responses) or aligning a tool with a stated objective (78% against 37%). The size of this difference is large by any benchmark, and the discussion treats its magnitude with corresponding caution.

RQ3 Pedagogical Readiness and Lesson Quality

Lesson-plan quality was measured only after the intervention, so the comparison for that outcome is between conditions at the post-test, and no improvement over time is claimed for it. Quality was higher in the experimental condition (M = 4.18, SD = 0.57) than in the control condition (M = 3.10, SD = 0.80), a difference of 1.08 points, 95% CI [0.78, 1.39], Welch t(74.2) = 7.14, p < .001, g = 1.54, 95% CI [1.06, 2.03]. Blinded expert ratings favoured the same condition (M = 4.46, SD = 0.44 against M = 3.00, SD = 0.70), a difference of 1.46, 95% CI [1.21, 1.71], Welch t(68.9) = 11.46, p < .001, g = 2.48, 95% CI [1.91, 3.05]. The expert evaluation, unlike lesson-plan quality, does have a baseline rating, and it was the same in the two conditions (M = 2.90, SD = 0.60). Over the term, that rating changed in the experimental condition (M change = 1.56, 95% CI [1.31, 1.81], t(41) = 12.45, p < .001) and did not change in the control condition (M change = 0.10, 95% CI [-0.20, 0.40], t(41) = 0.66, p = .512). The expert rubric separated the conditions on all four criteria: conceptual accuracy (92% against 78%), pedagogical appropriateness (88% against 65%), technological innovation (85% against 42%), and accessibility (76% against 39%). All four primary contrasts remain significant after the Holm correction.

The 168 archived lesson plans, one from each participant before and after the module, differed in kind and in degree. Control-condition plans leaned on static assets such as slides and PDF worksheets, whereas experimental-condition plans used dynamic 3D exploration inof the 42 plans, GIS habitat analysis in 72%, AR dissection in 64%, and AI identification in 57%. In 83% of experimental-condition plans, a tool choice was tied to a learning-design principle such as cognitive load or multimedia learning, against 27% in the control condition. Interaction logs from the final performance task, which both conditions completed in the same environment after the same introductory session, point the same way. Participants in the experimental condition manipulated tools more often (14.7 against 6.2 actions per session), annotated more (82% against 31% of sessions), and tested hypotheses more (68% against 21%).

The thematic analysis organised these differences into four themes. Under contextualization of technology, reviewers noted that experimental-group lessons connect a 3D model to a specific misconception about limb homology rather than showing it for its own sake. Under inquiry orientation, they described tasks that ask students to predict an outcome and then check it against the virtual specimen. Accessibility and conceptual precision separated the groups less sharply but in the same direction. Representative excerpts are given in Appendix D. Table 3 reports the coding frame in full, including the three counter-codes that record the absence of a practice.

Table 3

Coding frame for the qualitative strand: themes, codes, and segment frequencies by group

Theme / Code	Description	Segments	EG	CG
Contextualization of technology	Whether the digital tool is anchored in a specific biological idea or is present for its own sake.	96	63	33
Tool tied to a specific misconception	The plan names a known student error and introduces the tool at the point where that error arises, so that the representation does conceptual work the explanation alone cannot.	31	24	7
Representation matched to conceptual difficulty	The choice of representation follows from a property of the content, such as its three-dimensionality, its dynamics, or its spatial scale, and the plan states why a static image would not serve.	26	19	7
Tool used for display only (counter-code)	The tool appears without a learning task attached, and removing it would leave the lesson unchanged. Coded whenever the technology carries a presentational rather than an explanatory function.	22	7	15
Placement of the tool in the lesson sequence	The plan fixes when the tool enters, as a problem posed before the explanation, an illustration after it, or a check on understanding, and justifies that position.	17	13	4
Inquiry orientation	Whether the task positions the student as the one who asks and tests, or as the recipient of a demonstrated result.	88	58	30
Prediction before observation	Students commit to an expected outcome before the tool reveals it, so that the observation tests a claim instead of confirming a statement.	26	21	5
Student-generated question	The question under investigation is formulated by the student, and the tool is then used to pursue that question rather than a preset one.	22	16	6
Evidence used to revise an explanation	The task builds in a point at which the data obtained obliges the student to revise the initial explanation rather than record an expected result.	18	14	4
Confirmatory demonstration only (counter-code)	Work with the tool is limited to verifying what the instructor has already stated; the outcome is known in advance and no inference is required.	22	7	15
Accessibility	Whether the lesson is designed so that students with different equipment and preparation can reach the same objective.	66	38	28
Alternative route to the same task	The same objective is reachable by more than one path, so that a student who cannot use the primary representation still meets the content.	24	16	8
Adaptation for low-bandwidth or offline use	The plan anticipates unequal access to equipment and connectivity and provides a reduced-requirement version that preserves the biological task.	21	12	9
Graduated task difficulty	The task is tiered so that students entering with different levels of preparation work on the same biological content at an appropriate level of demand.	21	10	11
Conceptual precision	Whether the biological language accompanying the representation stays exact.	62	34	28
Structure distinguished from function	Annotations keep morphological structure separate from the function it serves, instead of naming a part by what it does.	23	15	8
Correct use of taxonomic terminology	Taxonomic rank, nomenclature, and morphological terms are used consistently and at the correct level of generality.	22	13	9
Ambiguous or conflated terminology (counter-code)	Terms are merged or used loosely, as when homology is treated as analogy, so that the representation transmits an imprecise concept.	17	6	11
Total (14 codes)		312	193	119

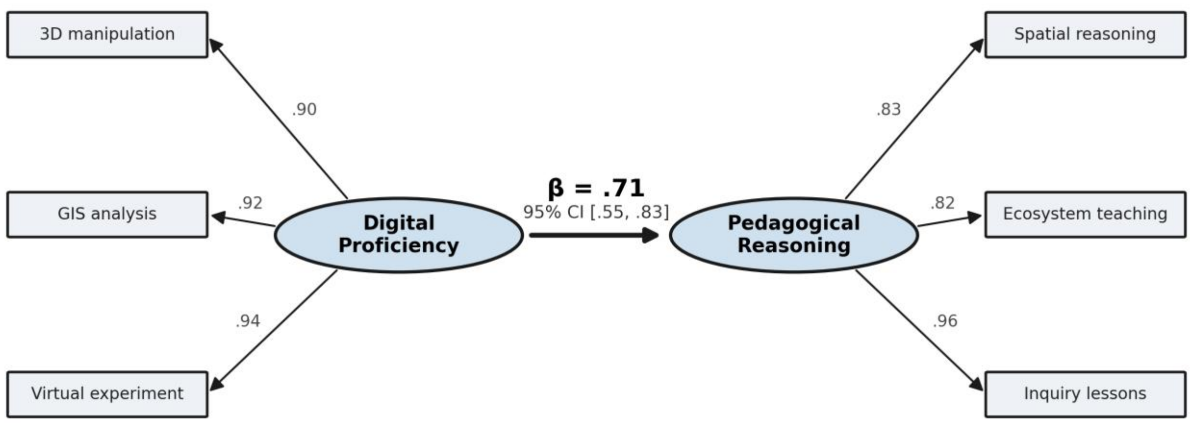
Note. The unit of analysis was the idea-bearing segment (N = 312). The corpus comprised the three expert evaluators' written comments, the 168 archived lesson plans (84 before and 84 after the module), and the virtual-laboratory interaction logs. Counter-codes record the absence of the practice a theme describes and are therefore expected to favor the control group. Intercoder reliability across the frame was Krippendorff's α = .82.

RQ4 Structural Relations between Tool Proficiency and Pedagogical Reasoning

This question is exploratory, and the model describes associations rather than effects. The measurement model fitted the full sample well: $\chi^2(8)$ = 10.41, p = .238, χ^2/df = 1.30, CFI = .994, TLI = .990, RMSEA = .060, 90% CI [.000, .150], SRMR = .027. The wide upper bound of the RMSEA interval follows from the modest sample and the small number of degrees of freedom. The structural coefficient from Digital Proficiency to Pedagogical Reasoning was positive and substantial, β = .71, with a percentile bootstrap interval of [.55, .83] over 2,000 resamples. Estimated separately within each condition, the same coefficient was .84 in the experimental condition and .65 in the control condition, and the three matched indicator pairs were associated in both, with r = .64, .61, and .72 in the experimental condition and r = .51, .58, and .51 in the control condition. These group-specific values are descriptive: measurement invariance was not tested, 42 cases per condition will not support a multi-group model, and no test of the difference between conditions was carried out. Because tool proficiency and pedagogical reasoning were measured at the same point in time, the coefficient describes how the two vary together and carries no implication about which precedes the other. Figure 2 reports the model together with the group-specific associations and their intervals.

Figure 2

Structural model of competency relations, with the matched indicator associations reported separately for each condition



Full-sample fit: $\chi^2(8)$ = 10.41, p = .238, χ^2/df = 1.30, CFI = .994, TLI = .990, RMSEA = .060 (90% CI .000–.150), SRMR = .027

Matched indicator associations by group (exploratory)

Indicator pair	Experimental (n = 42)	Control (n = 42)
3D manipulation and spatial reasoning	.64 [.33, .82]	.51 [.26, .70]
GIS proficiency and ecosystem teaching	.61 [.31, .79]	.58 [.37, .74]
virtual experiment and inquiry lessons	.72 [.52, .85]	.51 [.28, .70]
Structural coefficient, group-specific model	*.84*	*.65*

Pearson r with percentile bootstrap 95% intervals (2000 resamples). Group estimates are reported side by side and are not compared formally; measurement invariance was not tested and the sample does not support a multi-group model.

Discussion

The four research questions were answered in the same direction: the zoology-specific digital module was associated with larger gains than conventional teaching paired with a separate ICT course, and the difference appeared on every outcome rather than on one of them. The pattern is easier to interpret against the two traditions the module was built on. Shulman (1986) argued that teaching expertise is not subject knowledge and pedagogy held separately but their fusion in pedagogical content knowledge, and Mishra and Koehler (2006) extended that fusion to technology, describing technological pedagogical content knowledge as a complex and situated form of knowledge rather than an additive sum of three competences. Situated accounts of learning make the same point from the other side, holding that knowledge is in part a product of the activity and context in which it is acquired, so that skill learned apart from its domain of use tends to remain there (Brown et al., 1989). If those positions hold, digital competence taught outside a discipline should transfer poorly into disciplinary teaching, which is what the control condition allows us to examine. Each result is read against the relevant literature below, followed by what the study can and cannot support.

RQ1 Subject Knowledge

The larger gains in the experimental condition, concentrated in application items and in anatomical and ecological reasoning, are consistent with evidence that immersive and three-dimensional representations aid spatial and systems understanding in biology (Reinke et al., 2021; Sharmin et al., 2023; Tung & Chang, 2024). The distribution of the difference within the test matters more than its overall size. Recall items separated the conditions least, while items that required students to combine content with a teaching decision separated them most, which locates the difference in how knowledge was organised for use rather than in how much of it was retained. Studies of anatomy instruction with digital specimens report a similar asymmetry, the clearest benefits appearing where a structure has to be related to its spatial arrangement rather than named (Kavvadia et al., 2023; Pandey et al., 2024). Read through situated cognition; this is the expected outcome: the representation was tied to a task, and the knowledge acquired carried the shape of that task (Brown et al., 1989).

RQ2 Digital Didactic Competence

Gains across all three subscales, and especially in matching a tool to an objective, align with work arguing that TPACK develops when technological, pedagogical, and content demands are met together rather than in separate courses (Valtonen et al., 2020; Von Kotzebue, 2022). Reviewing 55 studies, Voogt et al. (2013) reached a conclusion that bears directly on the present design: accounts of TPACK within specific subject domains were almost absent from the literature, and active involvement in redesigning and teaching technology-enhanced lessons was among the few developmental strategies with evidence behind it. The module was built along that line, and the subscale on which the conditions diverged most, selecting a tool for a stated objective, is the one that cannot be exercised without holding content and pedagogy together. The result also speaks to reports that pre-service biology teachers leave generic ICT training able to operate software but unable to deploy it for their subject (Adhani & Nursia, 2024; Nurbaya, 2023), and to evidence that disciplinary framing and institutional support, rather than access to equipment, distinguish stronger from weaker pre-service profiles (Tondeur et al., 2017). Embedding the training appears to close part of that gap.

RQ3 Pedagogical Readiness and Lesson Quality

Higher lesson-quality and expert scores, together with the shift from static assets to dynamic, inquiry-orientated designs, match the situated-learning claim that professional skill forms through authentic practice rather than decontextualised drill (Singer et al., 2011). Synthesising 19 qualitative studies, Tondeur et al. (2012) identified authentic experience, reflection, feedback, and access to role models as the conditions under which pre-service teachers learn to integrate technology, and the weekly workshop cycle reproduced those conditions deliberately. The coded lesson plans locate the difference more precisely. What separated strong plans from weak ones was contextualisation, the connection of a tool to a specific conceptual difficulty, and this was also the code that discriminated most sharply between the conditions (Table 3). Heitink et al. (2017) describe the converse case in practice: teachers whose reasoning does not link a pedagogical strategy to the technology tend not to use the technology for that strategy at all, however available it is. The counter-codes in the present corpus behave the same way, with display-only tool use and confirmatory demonstration concentrated in the control condition. The baseline expert rating adds a detail worth keeping: it was the same in the two conditions and moved over the term only in the experimental one.

RQ4 Structural Relations

This is the most tentative part of the analysis and is reported as exploratory. Estimated on the full sample by maximum likelihood, the model satisfied the two-index criteria proposed by Hu and Bentler (1999), with CFI and TLI above their .95 threshold and an SRMR of .027 against their .08 benchmark; the measures entering the model and the estimation procedure are described in the Analytical Approach section, and their descriptive statistics appear in Table 2. Digital proficiency and pedagogical reasoning were strongly associated in the full sample, with a standardised coefficient of .71 and a bootstrap interval of [.55, .83]. Estimated within each condition the same coefficient was .84 in the experimental and .65 in the control condition, and the matched indicator pairs shown in Figure 2 were associated in both. Two conclusions follow, and they are worth separating. The association between tool proficiency and pedagogical reasoning is not confined to the experimental condition, so these data do not support the stronger claim that tool skill becomes pedagogically usable only when it has been acquired inside the discipline. What they do support is weaker and still substantive: the two constructs vary together wherever they were measured, and the level of both was higher where the preparation was discipline-embedded. This analysis does not test whether the two coefficients genuinely differ, for the reasons set out in the analytical approach, and they are reported side by side rather than contrasted. The pattern of association is consistent with Chai et al. (2013), who modelled the TPACK constructs in 550 pre-service teachers and found the basic knowledge factors acting on integrated knowledge indirectly, through intermediate combinations rather than by a direct route. One caution belongs with the fit statistics themselves. With eight degrees of freedom and 84 cases, the RMSEA confidence interval is wide by construction, and Kenny et al. (2015) show that RMSEA flags correctly specified models as poor fitting precisely in this range, so the point estimate of .060 should be read alongside the incremental indices rather than on its own. Because proficiency and reasoning were measured at the same time, none of this establishes which of them comes first.

Implications

For teacher educators the module offers a concrete route: identify the biological ideas that gain from visualisation, such as anatomy and ecological interaction, then build tool skill through tasks that teach those ideas. That students in the experimental condition were markedly better at aligning a tool with a learning goal–78% against 37% of optimal responses, points to the value of teaching digital-tool appraisal explicitly, a step conventional programmes often omit (Tondeur et al., 2012). The finding also bears on how competence frameworks are written. Instruments derived from DigCompEdu and comparable frameworks describe digital competence in general terms and are applied across subjects (Ghomi & Redecker, 2019; Tzafilkou et al., 2023), yet the capability that separated the conditions here is subject-specific, namely the judgement of which representation serves which biological idea. Voogt et al. (2013) noted the same absence of disciplinary specification in the research literature. These implications are addressed to course and programme design in teacher education, which is the level at which the study was conducted. A single site and 84 participants cannot support recommendations about certification or national competence standards, and none are offered here.

Context of Implementation

The study was carried out in a Kazakhstani teacher-education programme, and the constraints that shaped the module are shared across the region. Surveys of digital competence among teachers and pre-service teachers in Kazakhstan report uneven infrastructure and a reliance on general ICT provision rather than subject-specific preparation (Berkimbayev, 2021; Nazyrova et al., 2025; Zhukabayeva et al., 2025), and an assessment of future biology specialists in the same system found readiness for digital instruction lagging behind stated curricular expectations (Smagulova & Kalieva, 2025). The module was designed to operate under those conditions, which is why the accessibility strand of the coding frame recorded whether a lesson offered an alternative route to the same task or a reduced-requirement version (Table 3). That strand separated the conditions least, and the modest difference is itself informative: designing for unequal access was the aspect of digital didactics that the module changed least, and it is where a revised version would concentrate. The pattern is not confined to Kazakhstan. Analyzing the digitalization of the education system in the Kyrgyz Republic, Sulaimanova and Egamberdieva (2025) describe the same reliance on general strategies for raising teachers' digital competence rather than on subject-specific preparation.

Limitations

Limitations of Design and Inference

The strongest limitation is the allocation procedure. Two intact academic groups were assigned to the conditions, so the comparison is not protected by randomization, and although the standardized baseline differences were small and every interval included zero (Table 1), intervals of that width at 42 cases per condition cannot establish equivalence. Adjusting for the pre-test reduces but does not remove the possibility that an unmeasured difference between the two groups contributed to the outcome. No procedural barrier separated the conditions during the term, and although the elements of the module were not presented to students as anything special, exchange of materials between the groups cannot be excluded; such exchange would make the conditions more alike and would work against the difference reported here rather than inflate it. Both conditions were taught by the same instructor, who is also an author of this study, which holds instructor quality constant but leaves open an expectancy effect favouring the condition that the instructor designed. The module was new to the students who received it, so part of the difference may reflect that novelty rather than the design principle behind it. The intervention lasted one semester with no follow-up, so nothing reported here speaks to whether the difference persists or reaches classroom practice.

Limitations of Measurement and Magnitude

The effect sizes were large, and the largest of them, digital didactic competence at $d = 3.38$, lies far outside the range typical of educational interventions. Analysing 1,942 effect sizes from 747 randomised trials of education interventions, Kraft (2020) proposes .20 as the threshold for a large effect on standardised achievement outcomes, which places the present estimates an order of magnitude above the field's own benchmark. Part of that magnitude reflects the single-site design, the researcher-developed instruments, and the contrast with an untailored ICT control rather than with an equally intensive alternative, so the absolute sizes should be read as an upper estimate pending replication with independent measures. Some constructs relied on self-report. Reviewing 107 TPACK studies, Willermark (2018) found self-report to be the dominant mode of assessment and performance evaluation of actual teaching to be rare, which is why lesson-plan ratings, micro-teaching, blinded expert judgement, and interaction logs were included here, although the competence questionnaire itself remains a self-report measure. Blinding of the raters was incomplete in the way set out in the methodology section, since a plan that names a specific digital tool can betray the condition from which it came. The agreement between raters was computed on approximately one-fifth of the material and with unweighted coefficients; a weighted coefficient would suit an ordinal rubric better, and reporting one would have been preferable. The structural model, finally, rests on 84 cases and eight degrees of freedom, which is why it is presented as exploratory and why no comparison between conditions is drawn from it.

Future Research

The priorities that follow are longitudinal follow-up into early-career teaching, replication across biological subdisciplines and across institutions with different infrastructures, and lighter implementations of the module for settings without VR or AR equipment. A larger sample would also permit the multi-group model with measurement invariance that the present data cannot support, which is the only sound way to test whether the association between tool proficiency and pedagogical reasoning genuinely differs by condition, and it would give the model enough degrees of freedom for the fit indices to be informative in their own right (Kenny et al., 2015). A replication that sets the module against an equally intensive but subject-generic alternative would separate the contribution of disciplinary situating from that of additional instructional time, which the present design cannot fully disentangle. Where the institutional setting permits it, a design with random allocation would remove the main inferential limitation of the present study.

Conclusion

Embedding digital tools in the teaching of zoology, rather than teaching them alongside it, was associated with higher subject knowledge, digital didactic competence, and lesson quality among pre-service biology teachers, and tool proficiency and pedagogical reasoning varied together in both conditions. The contribution is not a new set of tools but a validated way of combining familiar ones into an assessed module, together with evidence, against an active control, that discipline-embedded digital preparation accompanies stronger teaching capability than a generic ICT course of the same length. Allocation was not random; the setting was one institution and one semester, and proficiency and reasoning were measured at the same point, so the immediate tasks are replication across settings, follow-up over time, and lighter implementations for resource-limited contexts. For teacher education the practical message is narrow: digital competence for biology teachers appears to be built more effectively through biology than beside it.

Acknowledgements

The authors thank Kamkin V. A. (Toraighyrov University, Pavlodar, Republic of Kazakhstan) for discussion of the study design and for comments on drafts of the manuscript.

Declaration of Interest

Authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The anonymised data supporting the findings of this study are available from the corresponding author on reasonable request. They are not publicly deposited, because the consent obtained from participants covered analysis and aggregate reporting by the research team rather than open publication of individual records.

Author Contributions

Zukhra Bagimbayeva – Conceptualization, Methodology, Investigation, Writing – original draft; Assiya Maimatayeva – Conceptualization, Methodology, Supervision; Zhadyra Kabatayeva – Investigation, Resources, Data curation; Kuralay Alipina – Investigation, Validation; Zhuldyz Tergenbayeva – Investigation, Project administration; Ruslan Safarov – Formal analysis, Software, Visualization, Writing – review and editing. All authors read and approved the final manuscript.

References

Abiltayeva, A., Zhumagulova, K., Satayev, M., Yechshzhanov, T., Maimatayeva, A., & Balta, N. (2025). Teacher competencies and digital integration into teaching practices: Perceptions from pre-service biology teachers in Kazakhstan. *Frontiers in Education*, *10*, Article 1628034. https://doi.org/10.3389/feduc.2025.1628034

Adhani, A., & Nursia, N. (2024). Pre-service biology teachers' digital competences profile. *Jurnal Pendidikan Indonesia Gemilang*, *4*(1), 54–66. https://doi.org/10.52889/jpig.v4i1.355

Ambarwati, R., Faizah, U., & Rahayu, D. A. (2019). Enhancing the digital literacy of pre-service biology teachers through an animal systematics course. In *Proceedings of the Mathematics, Informatics, Science, and Education International Conference (MISEIC 2019)*. Atlantis Press. https://doi.org/10.2991/miseic-19.2019.46

Anapana, G., Donkada, S., Surarapu, N., & Venkata Rathnamma, V. (2025). Multidisciplinary integration in zoology education: Bridging biological, computational, and environmental sciences. *Uttar Pradesh Journal of Zoology*, *46*(8), 317–331. https://doi.org/10.56557/upjoz/2025/v46i84916

Anniesmitha, K., Mathias, E. G., Bandekar, N. N., Rao, K. G. M., Aithal, P. A., & George, B. M. (2025). Visuospatial ability and neuroanatomy learning in health profession education – A scoping review. *Translational Research in Anatomy*, *41*, Article 100436. https://doi.org/10.1016/j.tria.2025.100436

Baker, K., Barnes, L., Rana, S., Jackson, J., & Kannan, R. (2023). Citizen science: Role of iNaturalist in biodiversity documentation and education in Arkansas. *Journal of the Arkansas Academy of Science*, *77*(1). https://doi.org/10.54119/jaas.2023.77119

Berkimbayev, K. M. (2021). The importance and models of forming the digital competence of teachers. *Bulletin of the Karaganda University Pedagogy Series*, *103*(3), 124–132. https://doi.org/10.31489/2021ped3/124-132

Biletska, H., Mironova, N., Yefremova, O., Barna, L., & Bloshchynskyi, I. (2021). The future biology teachers training for the implementation of sustainable development ideas in ecological education. *International Journal of Innovative Research and Scientific Studies*, *4*(4), 229–237. https://doi.org/10.53894/ijirss.v4i4.200

Boscolo-Berto, R., Tortorella, C., Porzionato, A., Stecco, C., Picardi, E. E. E., Macchi, V., & De Caro, R. (2021). The additional role of virtual to traditional dissection in teaching anatomy: A randomised controlled trial. *Surgical and Radiologic Anatomy*, *43*(4), 469–479. https://doi.org/10.1007/s00276-020-02551-2

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Chai, C. S., Ng, E. M. W., Li, W., Hong, H.-Y., & Koh, J. H. L. (2013). Validating and modelling technological pedagogical content knowledge framework among Asian preservice teachers. *Australasian Journal of Educational Technology*, *29*(1), 41–53. https://doi.org/10.14742/ajet.174

Chisingui, A. V., & Costa, N. (2020). Teacher education and sustainable development goals: A case study with future biology teachers in an Angolan higher education institution. *Sustainability*, *12*(8), Article 3344. https://doi.org/10.3390/su12083344

de Oliveira, G. A., Ferreira, M. B. C., Nunes, L. N., & Ribeiro, M. F. M. (2019). Pedagogical training profile of basic health sciences faculty in biomedical and related fields at Brazilian public and private higher education institutions. *Advances in Physiology Education*, *43*(2), 180–190. https://doi.org/10.1152/advan.00211.2018

Diem, H. T. T., Thinh, M. P., & Lam, V. T. T. (2024). Exploring practical pedagogy in high school biology education: A qualitative study of pre-service biology teachers' experiences in Vietnam. *European Journal of Educational Research*, *13*(2), 557–571. https://doi.org/10.12973/eu-jer.13.2.557

Dinçer, S. (2024). Bridging the gap in technology integration in education: An examination of science teachers’ competencies and needs. *Journal of Turkish Science Education*, *21*(4), 620–634. https://doi.org/10.36681/tused.2024.033

Duran, T., & Dikmenli, M. (2025). Use of artificial intelligence in biology education: A systematic review of literature. *Journal of Education in Science, Environment and Health*, *11*(4), 314–333. https://doi.org/10.55549/jeseh.864

Garzón-Artacho, E., Sola-Martínez, T., Romero-Rodríguez, J.-M., & Gómez-García, G. (2021). Teachers' perceptions of digital competence at the lifelong learning stage. *Heliyon*, *7*(7), Article e07513. https://doi.org/10.1016/j.heliyon.2021.e07513

Gaston, K. J., & O’Neill, M. A. (2004). Automated species identification: Why not? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1444), 655–667. https://doi.org/10.1098/rstb.2003.1442

Ghomi, M., & Redecker, C. (2019). Digital competence of educators (DigCompEdu): Development and evaluation of a self-assessment instrument for teachers' digital competence. In *Proceedings of the 11th International Conference on Computer Supported Education* (pp. 541–548). SCITEPRESS. https://doi.org/10.5220/0007679005410548

Hajiyeva, G. (2021). Teaching biology teaching methodology in higher pedagogical schools. In *Science, Education and Innovations in the Context of Modern Problems Conference*. IMCRA Publications.

Heitink, M., Voogt, J., Fisser, P., Verplanken, L., & van Braak, J. (2017). Eliciting teachers' technological pedagogical knowledge. *Australasian Journal of Educational Technology*, *33*(3), 96–109. https://doi.org/10.14742/ajet.3505

Herodotou, C., Ismail, N., Benavides Lahnstein, A. I., Aristeidou, M., Young, A. N., Johnson, R. F., Higgins, L. M., Ghadiri Khanaposhtani, M., Robinson, L. D., & Ballard, H. L. (2024). Young people in iNaturalist: A blended learning framework for biodiversity monitoring. *International Journal of Science Education, Part B*, *14*(2), 129–156. https://doi.org/10.1080/21548455.2023.2217472

Hsu, H.-P., Cheah, Y. H., & Hughes, J. E. (2023). A case study of a secondary biology teacher's pedagogical reasoning and action with augmented reality technology. *Education Sciences*, *13*(11), Article 1080. https://doi.org/10.3390/educsci13111080

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

International Society for Technology in Education. (2002). *National educational technology standards for teachers: Preparing teachers to use technology*. ISTE.

Irdalisa, I., Fuadi, T. M., Elvianasti, M., & Yanto, B. E. (2022). Technological pedagogical content knowledge: Ability prospective teachers biology education department in Jakarta Indonesia. *International Journal of Educational Research Review*, *7*(2), 114–123. https://doi.org/10.24331/ijere.1050594

Jensen, J. L. (2011). Higher education faculty versus high school teacher: Does pedagogical preparation make a difference? *Bioscene: Journal of College Biology Teaching*, *37*(2), 30–36.

Jiménez Sierra, Á. A., Ortega Iglesias, J. M., Cabero-Almenara, J., & Palacios-Rodríguez, A. (2023). Development of the teacher's technological pedagogical content knowledge (TPACK) from the lesson study: A systematic review. *Frontiers in Education*, *8*, Article 1078913. https://doi.org/10.3389/feduc.2023.1078913

Kang, N.-H., Yoo, J. E., & Kim, H. G. (2024). Development and validation of a scale of competency of digital age teaching (SCoDAT) for preservice teachers. *Innovation and Education*, *6*(1), 58–77. https://doi.org/10.1163/25248502-bja00005

Kavvadia, E.-M., Katsoula, I., Angelis, S., & Filippou, D. (2023). The Anatomage table: A promising alternative in anatomy education. *Cureus, 15*(8), Article e43047. https://doi.org/10.7759/cureus.43047

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kerski, J. J. (2019). Teaching demography and population change using web GIS tools and data. *The Geography Teacher*, *16*(3), 126–132. https://doi.org/10.1080/19338341.2019.1619608

Kim, J. Y., & Choi, J. K. (2025). Effects of AR on cognitive processes: An experimental study on object manipulation, eye-tracking, and behavior observation in design education. *Sensors*, *25*(6), Article 1882. https://doi.org/10.3390/s25061882

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Majewska, A. A., & Vereen, E. (2023). Using immersive virtual reality in an online biology course. *Journal for STEM Education Research*, *6*(3), 480–495. https://doi.org/10.1007/s41979-023-00095-9

Mašterová, V. (2023). Learning and teaching through inquiry with geospatial technologies: A systematic review. *European Journal of Geography*, *14*(3), 42–54. https://doi.org/10.48088/ejg.v.mas.14.3.042.054

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Nazyrova, A., Miłosz, M., Bekmanova, G., Omarbekova, A., Aimicheva, G., & Kadyr, Y. (2025). The digital transformation of higher education in the context of an AI-driven future. *Sustainability*, *17*(22), Article 9927. https://doi.org/10.3390/su17229927

Neumann, K., & Waight, N. (2020). The digitalization of science education: Déjà vu all over again? *Journal of Research in Science Teaching*, *57*(9), 1519–1528. https://doi.org/10.1002/tea.21668

Nietbaeva, F. (2023). Digital literacy in biology education through mobile learning technologies. *Eurasian Science Review*, *1*(4). https://doi.org/10.63034/esr-14

Nurbaya, N. (2023). Identifying prospective biology teachers' digital literacy competence at Cenderawasih University. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 10051–10058. https://doi.org/10.29303/jppipa.v9i11.4908

Ormandy, E., Schwab, J. C., Suiter, S., Green, N., Oakley, J., Osenkowski, P., & Sumner, C. (2022). Animal dissection vs. non-animal teaching methods. *The American Biology Teacher*, *84*(7), 399–404. https://doi.org/10.1525/abt.2022.84.7.399

Pandey, S. B., Bhadkaria, V., Chourasia, R. S., & Kumar, R. (2024). The role of technology in anatomy education: Evaluating knowledge gains with virtual dissection tools. *Journal of Heart Valve Disease*, *29*, 60–65. https://doi.org/10.47310/jhvd.2024.v29i01.011

Pererva, V. V., Lavrentieva, O. O., Lakomova, O. I., Zavalniuk, O. S., & Tolmachev, S. T. (2020). The technique of using a virtual learning environment in organizing future teachers' terminological work by specialty. *CEUR Workshop Proceedings, 2643*, 321–346. https://doi.org/10.31812/123456789/3868

Randler, C. (2008). Teaching species identification – A prerequisite for learning biodiversity and understanding ecology. *EURASIA Journal of Mathematics, Science and Technology Education*, *4*(3), 223–231. https://doi.org/10.12973/ejmste/75344

Reinke, N. B., Kynn, M., & Parkinson, A. L. (2021). Immersive 3D experience of osmosis improves learning outcomes of first-year cell biology students. *CBE—Life Sciences Education*, *20*(1), Article ar1. https://doi.org/10.1187/cbe.19-11-0254

Rode, Ž., & Torkar, G. (2023). The iNaturalist application in biology education: A systematic review. *International Journal of Educational Methodology*, *9*(4), 725–744. https://doi.org/10.12973/ijem.9.4.725

Sangur, K., Zubaidah, S., & Sulisetijono. (2025). A systematic literature review of mobile learning trends in biology education over ten years. *Social Sciences & Humanities Open*, *11*, Article 101429. https://doi.org/10.1016/j.ssaho.2025.101429

Sattar, M. A., Maqbool, M. U., Zakir, F., & Billah, M. (2025). Enhancing student engagement through augmented reality in secondary biology education. *Frontiers in Education*, *10*, Article 1628004. https://doi.org/10.3389/feduc.2025.1628004

Sautière, P. E., Blervacq, A. S., & Vizioli, J. (2019). Production and uses of e-learning tools for animal biology education at university. *The European Zoological Journal*, *86*(1), 63–78. https://doi.org/10.1080/24750263.2019.1582722

Sharmin, N., Chow, A. K., & King, S. (2023). Effect of teaching tools in spatial understanding in health science education: A systematic review. *Canadian Medical Education Journal, 14*(4), 70–88. https://doi.org/10.36834/cmej.74978

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Singer, J., Lotter, C., Feller, R., & Gates, H. (2011). Exploring a model of situated professional development: Impact on classroom practice. *Journal of Science Teacher Education*, *22*(3), 203–227. https://doi.org/10.1007/s10972-011-9229-0

Smagulova, A. E., & Kalieva, A. N. (2025). Evaluating the preparedness of future biology specialists to integrate STEM approaches into their education and training. *Bulletin of Ablai Khan KazUIRandWL, Series "Pedagogical Sciences"*, *77*(2), 65–79. https://doi.org/10.48371/PEDS.2025.77.2.004

Sulaimanova, R., & Egamberdieva, A. (2025). Professional development of teachers in the context of digital transformation. *Bulletin of the Jusup Balasagyn Kyrgyz National University*, 9–20. https://doi.org/10.58649/1694-8033-2025-3(123)-9-20

Tanner, K., & Allen, D. (2006). Approaches to biology teaching and learning: On integrating pedagogical training into the graduate experiences of future science faculty. *CBE—Life Sciences Education*, *5*(1), 1–6. https://doi.org/10.1187/cbe.05-12-0132

Telecan, T., Capraș, R.-D., Filip, G. A., Bonea, M., & Crivii, C.-B. (2025). Dissection in the 21st century: Virtual tables versus traditional methods and their influence on medical students' perception – A systematic review. *BMC Medical Education*, *25*(1), Article 1332. https://doi.org/10.1186/s12909-025-07946-6

Tergenbayeva, Z., Karasholakova, L., Çatar Sarıkurt, B., Özdilek, Z., Atasoy, E., & Kitapbayeva, A. (2023). Investigation of pre-service science teachers' knowledge levels, practical experiences, and perceived competencies in teaching biodiversity. *Journal of Turkish Science Education*, *20*(4), 695–717. https://doi.org/10.36681/tused.2023.039

Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, *59*(1), 134–144. https://doi.org/10.1016/j.compedu.2011.10.009

Tondeur, J., Scherer, R., Siddiq, F., & Baran, E. (2017). A comprehensive investigation of TPACK within pre-service teachers' ICT profiles: Mind the gap! *Australasian Journal of Educational Technology*, *33*(3), 46–60. https://doi.org/10.14742/ajet.3504

Tung, Y.-H., & Chang, C.-Y. (2024). How three-dimensional sketching environments affect spatial thinking: A functional magnetic resonance imaging study of virtual reality. *PLOS ONE*, *19*(3), Article e0294451. https://doi.org/10.1371/journal.pone.0294451

Tzafilkou, K., Perifanou, M., & Economides, A. A. (2023). Assessing teachers' digital competence in primary and secondary education: Applying a new instrument to integrate pedagogical and professional elements for digital education. *Education and Information Technologies*, *28*(12), 16017–16040. https://doi.org/10.1007/s10639-023-11848-9

Väätäjä, J. O., & Ruokamo, H. (2021). Conceptualizing dimensions and a model for digital pedagogy. *Journal of Pacific Rim Psychology*, *15*, Article 1834490921995395. https://doi.org/10.1177/1834490921995395

Valtonen, T., Leppänen, U., Hyypiä, M., Sointu, E., Smits, A., & Tondeur, J. (2020). Fresh perspectives on TPACK: Pre-service teachers' own appraisal of their challenging and confident TPACK areas. *Education and Information Technologies*, *25*(4), 2823–2842. https://doi.org/10.1007/s10639-019-10092-4

von Kotzebue, L. (2022). Beliefs, self-reported or performance-assessed TPACK: What can predict the quality of technology-enhanced biology lesson plans? *Journal of Science Education and Technology*, *31*(5), 570–582. https://doi.org/10.1007/s10956-022-09974-z

Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge – a review of the literature. *Journal of Computer Assisted Learning*, *29*(2), 109–121. https://doi.org/10.1111/j.1365-2729.2012.00487.x

Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, *9*(11), 2216–2225. https://doi.org/10.1111/2041-210X.13075

Willermark, S. (2018). Technological pedagogical and content knowledge: A review of empirical studies published from 2011 to 2016. *Journal of Educational Computing Research*, *56*(3), 315–343. https://doi.org/10.1177/0735633117713114

Zhang, J., Su, T., Liang, X., Xu, Y., Wang, Z., Yu, Y., & Ge, J. (2023). The mediating effect of geospatial thinking on the relationship between family capital and academic achievement in geography. *Frontiers in Psychology*, *14*, Article 1067198. https://doi.org/10.3389/fpsyg.2023.1067198

Zhukabayeva, T., Baumuratova, D., Zholshiyeva, L., Karabay, A., & Abdrakhmanov, K. (2025). Digital transformation in higher education: Toward a national model of digital university in Kazakhstan through global and local comparison. *Sustainability*, *17*(24), Article 11132. https://doi.org/10.3390/su172411132

Appendix

Appendix A

Comparative Thematic Plan of the "Zoology" Course (Fragment)

Week / Module	Topic	Control Group (CG) Activities	Experimental Group (EG) Activities	Digital Tools (EG only)
Week 1	Introduction to Zoology	• Traditional lecture with static Power Point presentation. • Laboratory work: Tinbergen's Four Questions (Mechanism, Development, Function, Evolution). • Board game "Lotto", creating an animal "passport" by filling in all the taxa on cardboard cards.	• Interactive lecture with 3D visualization. • Laboratory work: Tinbergen's Four Questions (Mechanism, Development, Function, Evolution). • Functional modeling of fundamentals of animal taxonomy on laptops.	• Animal Diversity Web (ADW) (Data source) • OneZoom Tree of Life Explorer (Phylogenetic visualization) • iTOL (Interactive Tree Of Life)
Weeks 2-7	Invertebrate zoology	• Examination of museum collections (pinned insects). • Identification using dichotomous paper keys. • Rote memorisation of orders/families.	• "Digital collection" creation. • Field identification using AI-driven image recognition. • Creating a digital phylogenetic tree.	• iNaturalist / Seek • Google Lens • iTOL (Interactive Tree Of Life)
Weeks 8-13	Vertebrate zoology	• Traditional lecture using static diagrams. • Laboratory work: Dissection of preserved wet specimens (frogs). • Sketching organ systems in notebooks.	• Interactive lecture with 3D visualization. • Laboratory work: Virtual dissection with "undo" function and layer isolation. • Functional modeling of the circulatory system.	• Biosphera 3D (Comparative anatomy: Fish, Frog, Cat modules) • Froggipedia (AR interaction) • Emantras Virtual Frog Dissection (Skill training)
Week 14-15	Ecology & Biogeography	• Textbook study of habitat maps. • Analysis of static climate charts. • Discussion on local fauna.	• Geospatial analysis of species distribution. • Overlaying climate data on migration routes. • Project: "Mapping local biodiversity loss".	• QGIS / Google Earth Pro • DIVA-GIS • GBIF Database • IUCN Red List
Week 16	Final Assessment	• Theory: Standard written exam. • Performance Task: Solving a diagnostic biological problem using a Virtual Laboratory (without prior training, to measure digital intuition and baseline skill). • Oral defense of laboratory notebook.	• Theory: Standard written exam. • Performance Task: Solving a complex biological problem using a Virtual Laboratory (demonstrating acquired skills). • Defense of digital portfolio.	• LMS (Moodle) • Presentation Software • Virtual Dissection / Simulation Software (Both groups)

Appendix B

Sample Worksheets for the Four Tool Families

Each worksheet occupied one weekly workshop. Only the core task and the pedagogical-reflection prompt are reproduced here; full materials are available from the authors.

B1. Virtual Dissection (Frog Urogenital System)

Task: locate and label the kidneys, ureters, and gonads on the virtual specimen; record two structures you initially mis-identified and the cue that corrected you. Reflection: which feature of the virtual tool (undo, rotation, layer isolation) most helped a student who struggles with 3D relationships, and why?

B2. 3D / AR Morphology (Vertebrate Limb Homology)

Task: rotate and cross-section the forelimb of three vertebrates; mark the homologous bones and note where a 2D diagram would hide the relationship. Reflection: design a two-minute prompt that uses the model to surface, then correct, the common misconception that analogous limbs are homologous.

B3. GIS Species-Distribution Mapping (Local Passerine)

Task: import occurrence points, overlay a land-cover layer, and describe one pattern of habitat association and one likely sampling bias. Reflection: turn the map into an inquiry question that a school class could investigate with free data.

B4. AI Species Identification (Field Session)

Task: photograph and identify five organisms with an AI app; for each, record the top suggestion, your confidence, and the diagnostic feature you used to confirm or reject it. Reflection: where did the AI fail, and how would you use that failure as a teaching moment about taxonomic characters?

Appendix C

Scoring Rubrics and Sample Items

C1. Analytic Rubric for Open-Ended Subject-Knowledge Items (0–4 Per Item)

Level	Descriptor
4 (full)	Biological content correct and complete, and explicitly translated into a teaching move.
3 (proficient)	Content correct; teaching translation present but partial.
2 (developing)	Content largely correct; little or no pedagogical translation.
1 (limited)	Fragmentary content, or a serious conceptual error.
0	No creditable response.

C2. Sample Subject-Knowledge Items

MCQ (anatomy): In the amphibian heart, deoxygenated and oxygenated blood mix in the (a) left atrium; (b) right atrium; (c) single ventricle; (d) truncus arteriosus. [key: c]

Open-ended (application): A student claims the bat wing and the insect wing are homologous because both are used for flight. Outline, in three sentences, how you would use a 3D model to correct this. [scored with rubric C1]

C3. Digital Didactic Competence Questionnaire – Sample Items by Subscale (5-Point Likert, 1 = Strongly Disagree to 5 = Strongly Agree, Unless Noted)

Tool selection: "Given a zoology topic, I can choose the digital tool whose representation best fits the learning goal." Pedagogical rationale (open, scored 0–3): "Explain why you would use a GIS map rather than a static distribution figure for teaching habitat fragmentation." Technical troubleshooting (scenario): "A VR headset fails to load a model mid-lesson; state your first two actions."

C4. Analytic Rubric For Lesson Plans And Micro-Teaching (Each Criterion 1–5).

Criterion	1–2	3	4–5
Conceptual accuracy	Errors present	Mostly accurate	Accurate and precise
Pedagogical appropriateness	Tool-led, goal unclear	Goal present	Tool serves an explicit goal
Technological integration	Decorative use	Functional use	Tool addresses a specific difficulty
Accessibility	Not considered	Partially considered	Differentiation designed in

Note. Inter-rater reliability on a common subset: ICC = .89 (open-ended items); κ = .84 (lesson-plan rubric); κ = .81 (micro-teaching rubric).

Appendix D

Representative Qualitative Excerpts by Theme

Excerpts are drawn from the three experts' written comments on experimental-group lessons; each is labelled by the theme it illustrates.

Contextualization of Technology

"The 3D limb model is not shown for spectacle; it is used at the exact point where students confuse homology with analogy." / "The GIS layer is tied to a question the class can actually test."

Inquiry Orientation

"Students are asked to predict the internal arrangement before the virtual dissection reveals it, which turns the tool into a test of reasoning rather than a display."

Accessibility

"The plan offers the same task through the model and through a labelled still, so learners who tire in VR are not excluded."

Conceptual Precision

"Terminology is used carefully; the annotation distinguishes a structure from its function instead of collapsing the two."