From Tool Literacy to Workflow Fluency: Designing AI-agent Exemplary Cases for the Professional Development of Early-career University Faculty

Ping Zhang*, Juliet Zhu
School of Computer Science, Baoji University of Arts and Sciences, Baoji 721016, China
*Corresponding email: cszhangping@bjwlxy.edu.cn

Abstract

Most institutional training for early-career university teachers still treats artificial intelligence (AI) as a set of stand-alone tools to be demonstrated in one-off seminars. Teachers leave such sessions able to name a chatbot but unable to fold it into the messy, multi-step tasks that fill a working week. We report on the design, delivery, and evaluation of a different approach: a layered, scenario-based training pathway organised around reusable AI-agent exemplary cases. The cases were built on an accessible low-code stack (Coze for workflow orchestration, retrieval-augmented generation flow for retrieval-augmented question answering over local corpora, a fine-tuned Llama 3 model for domain question answering, and robotic process automation for routine office work), and were routed to participants according to a diagnostic competency profile. Using a quasi-experimental design at a regional comprehensive university in western China, we compared an experimental group that received the case-based pathway (n=45) with a comparison group that received a conventional lecture-style AI-literacy workshop (n=41) over one semester. Self-rated competency across six dimensions, task-completion time on four authentic tasks, and a technology-acceptance survey were collected, complemented by twelve semi-structured interviews. The experimental group improved substantially on every competency dimension, with the largest gains in agent and workflow construction; between-group differences on the post-test were significant after controlling for the pre-test (ANCOVA, $p<0.001$), and the time needed for four representative tasks fell by 55-78%. Acceptance ratings were high, and interviews pointed to a shift from isolated tool use toward thinking in workflows. We argue that exemplary cases, paired with diagnostic routing, offer a transferable template for institutions seeking to move faculty AI training beyond tool demonstrations.

Keywords

Artificial intelligence agents, Faculty professional development, Early-career teachers, Large language models, Retrieval-augmented generation, Low-code workflow, Mixed-methods

Introduction

The rapid diffusion of large language models has changed what it means to be a competent university teacher. A lecturer is now expected not only to know discipline but also to draft course materials, summaries a fast-moving literature, analyses student data, and clear a steady stream of administrative correspondence, increasingly with the help of generative tools. Early-career faculty feel this pressure most acutely. They carry heavy teaching loads, are still establishing a research line, and rarely have the slack to teach themselves a new technology by trial and error. China's national policies have clarified relevant requirements. Currently, efforts are being intensified to promote the digital transformation of education. Accordingly, educational institutions are urged to enhance teachers' digital and AI competencies. Such competency development has become a fundamental obligation instead of a discretionary add-on.

Yet the way most universities respond to this call has not kept pace with the tools themselves. The dominant format remains the one-off seminar: An expert demonstrates a chatbot or a slide-generator to a roomful of teachers, who applaud, take a few screenshots, and return to offices where almost nothing changes. We have run and observed enough of these sessions to recognize the pattern three weaknesses arose. Training tends to be homogeneous, delivering the same content to a chemist and a historian regardless of their starting point. It is fragmented, presenting tools in isolation rather than as parts of a task. It is disconnected from practice, so that the gap between the polished demonstration and the teacher's own half-formed problem is left for the teacher to bridge alone. The result is a familiar disappointment. Teachers can recite the names of tools but cannot assemble them into anything that survives contact with a real Monday morning.

Part of the problem is conceptual. Most existing training programs for faculty focus merely on tool literacy. Tool literacy refers to the ability to operate specific software applications. Nevertheless, a more essential skill is what we term workflow fluency. Workflow fluency enables educators to break down practical tasks. It also helps them select appropriate tools and formulate operational steps. Furthermore, practitioners can combine these elements into repeatable and shareable working processes. A teacher who has workflow fluency does not ask whether a model can write an email; she has a small, dependable pipeline that turns a folder of notices into ready-to-send drafts and trusts it enough to stop thinking about it. Reaching that state is not a matter of seeing more demonstrations. It requires working through concrete, transferable cases that are close enough to one's own job to be adapted rather than admired.

This study grew out of an attempt to redesign faculty AI training around that idea. Working at a regional comprehensive university, we developed a set of AI-agent exemplary cases on a deliberately accessible low-code stack and organized them into a layered pathway that meets teachers at their diagnosed level rather than at a single notional average. An exemplary case, in our usage, is a documented, runnable solution to an authentic faculty task, packaged so that a teacher can reproduce it, inspect how it works, and modify it for a neighboring task. The cases span teaching, research, and administrative scenarios, and they are routed to participants by a lightweight competency profile so that a novice is not handed an advanced fine-tuning exercise on day one.

We pursued three questions. First, does a layered, case-based pathway raise early-career teachers' self-rated AI competence more than a conventional workshop, and across which dimensions? Second, does working with the agent cases translate into measurable efficiency gains on authentic tasks rather than merely improving confidence? Third, how do participants experience the shift, and what conditions appear necessary for it to take hold? To answer these, we ran a one-semester quasi-experiment with a comparison group, combining competency and efficiency measures with a technology-acceptance survey, and interviewed a subset of participants. The contribution is not a new model architecture but a training framework: A transferable template grounded in evidence for shifting institutional faculty AI training from tool demonstration to workflow practice.

Background and related work

AI in faculty professional development

Faculty development has long been framed through models such as the Technological Pedagogical Content Knowledge framework, which insists that technology competence is meaningful only when interwoven with pedagogy and subject matter [1]. The arrival of generative AI has prompted a wave of work extending this view to AI literacy, typically describing a progression from awareness, through hands-on use, to critical and ethical judgement. Reviews of AI in higher-education teaching converge on a common finding: Enthusiasm is high, but sustained behavior change is rare, and training that stops at awareness seldom produces it. Literature increasingly distinguishes between knowing that a tool exists and being able to integrate it into routine professional work, which is precisely the distinction our design tries to operate.

From single tools to AI agents and workflows

Early generative-AI training, in universities and elsewhere, concentrated on prompting a single chatbot. The field has since moved toward agents: Systems can plan a sequence of steps, call external tools, retrieve information, and act with a degree of autonomy. Low-code orchestration platforms such as Coze have made it possible to assemble such agents without writing much code, lowering the barrier for non-specialists. This shift matters for training because it changes the unit of competence. The relevant skill is no longer formulating one clever prompt but composing a workflow, an idea that resonates with long-standing arguments in computational-thinking education about decomposition and abstraction. Surprisingly little published work, however, examines how to teach workflow composition to working academics as opposed to students.

Retrieval-augmented generation (RAG) is central to making such workflows trustworthy in an academic setting [2]. By grounding a model's output in a curated local corpus, RAG reduces fabrication and lets teachers query their own course materials, regulations, or research notes. Open implementations such as retrieval-augmented generation flow (RAG Flow) have made local, privacy-preserving knowledge bases feasible on modest hardware, which is an important consideration for teachers handling unpublished or sensitive material. Parameter-efficient fine-tuning of open models such as Llama 3 further allows a department to adapt a model to its own terminology without prohibitive cost. Robotic process automation, finally, addresses the unglamorous but time-consuming administrative load that crowds out teaching and research.

The gap

Two gaps motivate this study. The first is the scarcity of evidence on training designs that treat AI as composable workflows rather than discrete tools; most reported interventions still evaluate awareness or attitude rather than transferable capability. The second is the near absence of personalized, layered approaches: Training is overwhelmingly delivered as one course to a heterogeneous audience, despite wide variation in teachers' starting competence and disciplinary needs. The exemplary case pathway we describe is an attempt to address both at once, and to do so with a cheap and simple enough that other institutions could plausibly reuse it.

Methods

Research design

We adopted a mixed methods, quasi-experimental design with a non-equivalent comparison group, embedded within an action research cycle of diagnosis, intervention, evaluation, and refinement (Lewin's tradition of iterative practical inquiry). A randomized trial was neither feasible nor, we judged, appropriate: Participation in faculty development is voluntary, and withholding a promising intervention by lottery would have been hard to justify. Group assignment therefore followed the training cohort teachers had enrolled in, and we used analysis of covariance to adjust for pre-existing differences. Quantitative measures answered the first two research questions; interviews answered the third and helped interpret the numbers.

Participants and setting

The study took place over the autumn semester of 2025 at a regional comprehensive university in western China. Participants were early-career faculty, defined as teachers aged 40 years old or under with no more than five years of service. A total of 92 teachers initially joined the program, and 86 completed all measures (retention rate: 93%); the six who dropped out cited workload and were not analyzed. The experimental group (n=45) followed the case-based pathway; the comparison group (n=41) attended a conventional three-session AI-literacy workshop covering the same nominal topics through lectures and live demonstrations. Table 1 summarizes the two groups, which were comparable to the observed background variables.

Table 1. Participant characteristics by group.

Characteristic	Experimental (n=45)	Comparison (n=41)	p
Mean age, years (SD)	32.4 (3.60)	33.1 (3.90)	0.39
Female, n (%)	26.0 (57.80)	22.0 (53.70)	0.70
Holds doctorate, n (%)	29.0 (64.40)	25.0 (61.00)	0.74
Mean years of service (SD)	2.8(1.40)	3.0 (1.50)	0.52
STEM discipline, n (%)	24.0 (53.30)	20.0 (48.80)	0.67
Prior structured AI training, n (%)	11.0 (24.40)	9.0 (22.00)	0.79

Note: Group differences were tested with independent-samples t-tests for continuous variables and chi-square tests for categorical variables; none approached significance.

The training framework

The intervention rests on a four-layer framework (Figure 1). At the top sit the three scenario domains that organise a teacher's working life: teaching, research, and administration. Beneath them is the agent layer, where concrete capability is assembled from Coze workflows, RAG-based question answering, robotic process automation (RPA), and, for advanced users, multi-agent orchestration. A personalised adaptation engine links the two: It takes a teacher's diagnostic competency profile and routes her to cases at an appropriate level of difficulty. The model and data layer supplies the underlying machinery, a fine-tuned Llama 3 model, local knowledge bases, and data-mining and natural-language-processing components, while the infrastructure layer provides the low-code platform, computing resources, connectors, and the access and ethics controls that keep sensitive material in safe hands.

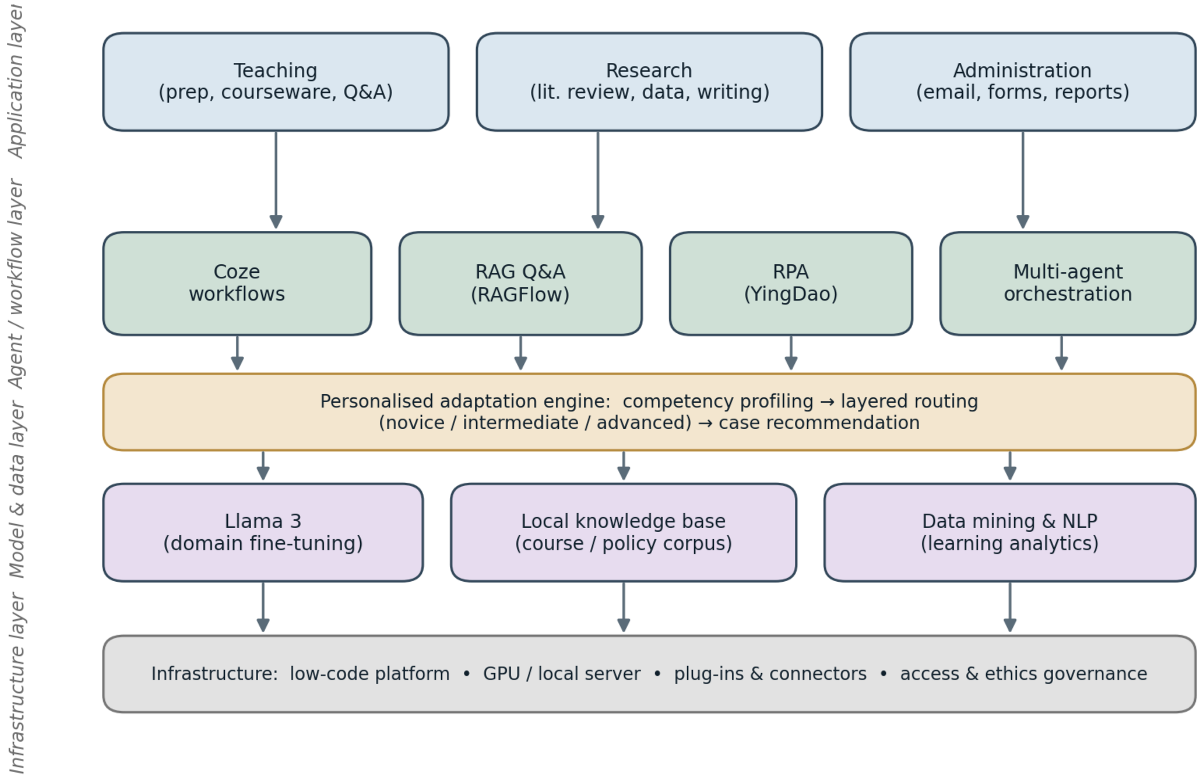

Figure 1. Four-layer architecture of the AI-agent faculty-training framework.

The choice of stack was deliberate. We favored low-code and open components over proprietary all-in-one platforms because the project's aim was reproducibility on a small budget, and because teachers who can see inside a workflow are more likely to adapt it than those handed a sealed product.

Exemplary case design

Six case modules form the spine of the pathway (Table 2). Each module is an authentic task rendered as a runnable, documented solution, with a short rationale, a step-by-step build, a sample input and output, and a set of suggested variations the teacher can attempt. The modules are ordered by difficulty, beginning with a single-step Coze summarization workflow and ending with a multi-agent process that chains several of the earlier components. Crucially, the cases are not discipline-specific in their mechanics but are presented with discipline-varied examples, so that the historian and the chemist each see the summarizations case applied to material they recognize.

Table 2. The six exemplary-case modules.

ID	Module	Core technology	Representative task
M1	Literature & news summarizations workflow	Coze	Condense a batch of articles or papers into a structured brief
M2	Local knowledge-base Q&A	RAGFlow (RAG)	Answer questions over course materials and institutional regulations

ID	Module	Core technology	Representative task
M3	Domain question answering / grading aid	Llama 3 fine-tuning	Respond in field terminology; draft feedback on short answers
M4	Office automation	RPA (YingDao)	Turn a folder of notices into batch email and form drafts
M5	Learning analytics	Data mining & NLP	Cluster student performance and flag at-risk patterns
M6	Multi-agent orchestration	Integration of M1-M5	Chain retrieval, drafting and dispatch into one reusable pipeline

The implementation pathway

Delivery followed four phases across the semester. In the diagnosis phase (weeks 1-2) every participant completed the competency self-assessment and a short skills task, producing the profile used for routing. The co-development phase (weeks 3-7) built and refined the case modules with a small group of volunteer teachers, ensuring the examples felt real rather than contrived. The pilot phase (weeks 8-13) delivered the cases, with each teacher entering at her diagnosed level (novice, intermediate, or advanced) and progressing as she demonstrated mastery; a shared chat channel and weekly office hours provided support. The evaluation phase (weeks 14-16) collected post-measures, ran interviews, and folded the findings back into a revised case library and an implementation handbook. The comparison group received its three workshop sessions during weeks 8-13 to keep the calendar comparable.

Instruments and data collection

Competency was measured with a six-dimension self-report instrument covering foundational tool operation, agent and workflow construction, teaching application, research application, office automation, and ethical and critical use. Each dimension used four items on a five-point Likert scale; internal consistency was acceptable to good (Cronbach's alpha between 0.79 and 0.88 across dimensions). Task efficiency was measured directly: Each participant completed four authentic tasks under timed conditions before training and again after, once without and once with agent assistance, and we recorded completion time. Acceptance was assessed with an eight-item survey adapted from the Technology Acceptance Model and the unified theory of acceptance and use of technology, yielding scores for perceived usefulness, perceived ease of use, satisfaction, and continuance intention [3,4]. Finally, twelve participants from the experimental group, sampled to vary by discipline and entry level, took part in semi-structured interviews of roughly thirty minutes.

Data analysis

Quantitative data was analyzed in standard statistical software. Within-group change was tested with paired-samples t-tests, between-group post-test differences were tested with analysis of covariance using the pre-test as covariate, reporting partial eta squared as the effect-size measure. Efficiency data were summarized as mean completion times and percentage reductions. Interview transcripts were analyzed with reflexive thematic analysis: Two researchers coded independently, met to reconcile, and developed themes iteratively [5,6]. We treated the qualitative strand as a way to explain and qualify the quantitative results rather than as a separate confirmatory test.

Results

Competency gains

Both groups improved over the semester, but the scale of improvement differed sharply. Figure 2 shows the six-dimension competency profile for each group before and after training. The experimental group expanded on every axis, with the most pronounced movement on agent/workflow construction, the dimension that conventional training tends to neglect. The comparison group improved modestly and unevenly, gaining most on foundational operation and least on the construction and application dimensions.

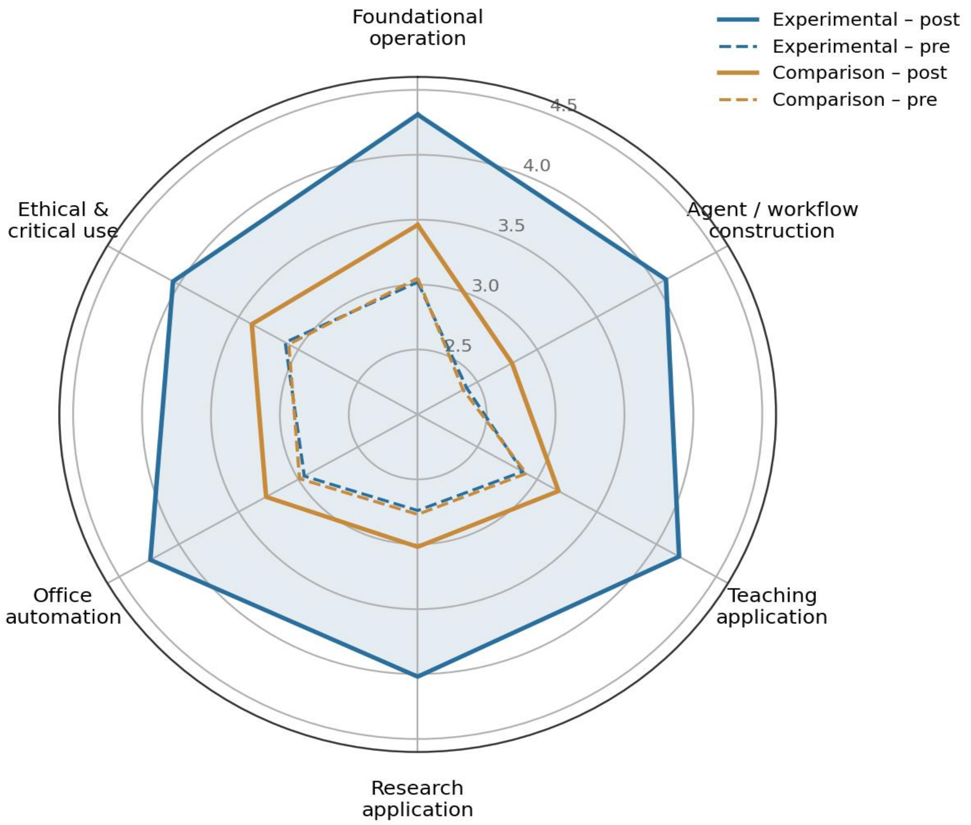

Figure 2. Self-rated competency across six dimensions, by group, pre and post.

Table 3 reports the figures behind the chart. Within the experimental group, every dimension rose by more than a full-scale point, with large effect sizes throughout. After adjusting for the pre-test, between-group differences on the post-test were significant for all six dimensions and for the composite score (ANCOVA, $p<0.001$ in every case), with partial eta squared values indicating that group membership accounted for roughly half the variance in adjusted post-test scores. The pattern is consistent with the design's emphasis: The cases that asked teachers to build and combine workflows produced the gains that mattered most.

Table 3. Pre-training and post-training competency by dimension (five-point scale; mean (SD)).

Dimension	Exp. pre	Exp. post	Cmp. pre	Cmp. post	d (exp)	p
Foundational operation	3.02 (0.61)	4.31 (0.48)	3.05 (0.63)	3.46 (0.59)	2.36	<0.001
Agent / workflow construction	2.41 (0.70)	4.08 (0.55)	2.38 (0.68)	2.79 (0.64)	2.64	<0.001
Teaching application	2.88 (0.66)	4.19 (0.51)	2.91 (0.64)	3.18 (0.60)	2.22	<0.001
Research application	2.74 (0.72)	4.02 (0.57)	2.77 (0.70)	3.02 (0.66)	1.97	<0.001
Office automation	2.95 (0.64)	4.24 (0.49)	2.99 (0.66)	3.27 (0.62)	2.27	<0.001
Ethical & critical use	3.11 (0.69)	4.05 (0.58)	3.08 (0.67)	3.39 (0.63)	1.47	<0.001
Composite	2.85 (0.52)	4.15 (0.41)	2.86 (0.53)	3.19 (0.49)	2.78	<0.001

Task efficiency

The competency gains were mirrored in how long authentic tasks actually took. Figure 3 contrasts mean completion times before and after agent assistance on the four timed tasks. The reductions were large and consistent: 73% for literature/news summarisation, 60% for courseware drafting, 55% for a data-analysis report, and 78% for batch administrative correspondence. The two tasks with the steepest savings, summarisation and routine correspondence, are exactly the kind of repetitive work that early-career teachers had described as crowding out their core duties, which may explain why participants spoke about the office-automation case with particular warmth.

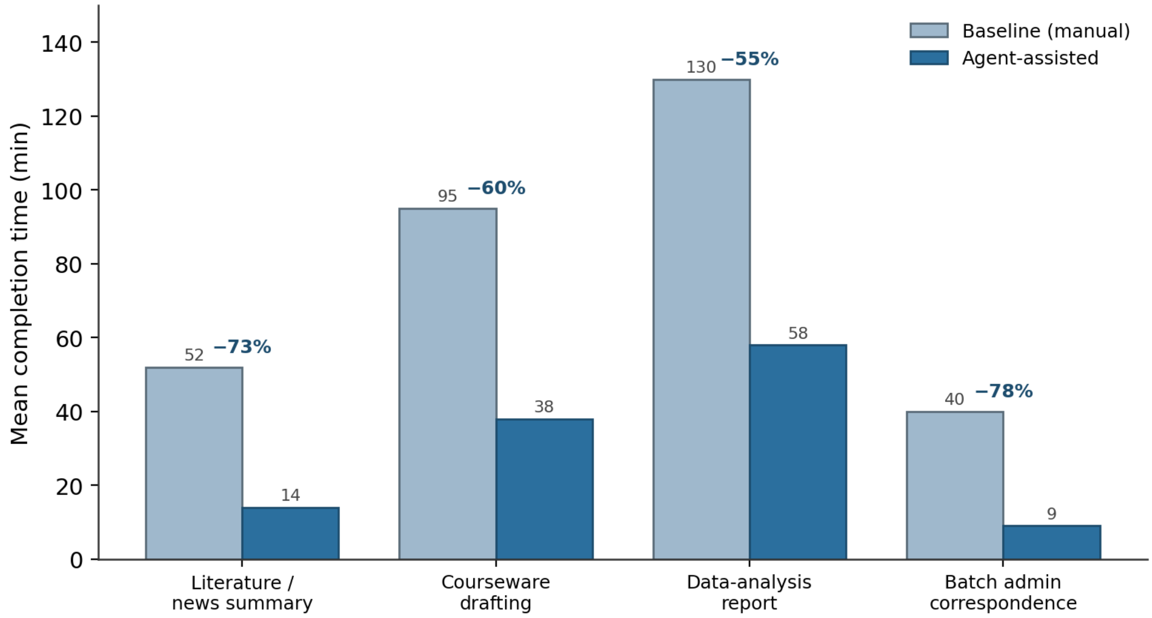

Figure 3. Mean task-completion time before and after agent assistance (experimental group).

We note that the slowest task to speed up, the data-analysis report, was also the one requiring the most judgement, and several teachers chose to keep a larger share of the work manual. The agents saved time without, in their telling, taking over the thinking.

Acceptance

Acceptance of the case-based training was high across all four constructs (Figure 4). Perceived usefulness drew the strongest endorsement (M=4.32, SD=0.47 on a five-point scale), followed by continuance intention (M=4.27) and satisfaction (M=4.21); perceived ease of use, while still well above the neutral point, was the lowest of the four (M=4.06, SD=0.55). That ordering is itself informative. Teachers were readily convinced the cases were useful and intended to keep using them, but a minority still found parts of the build process demanding, a tension we return to below.

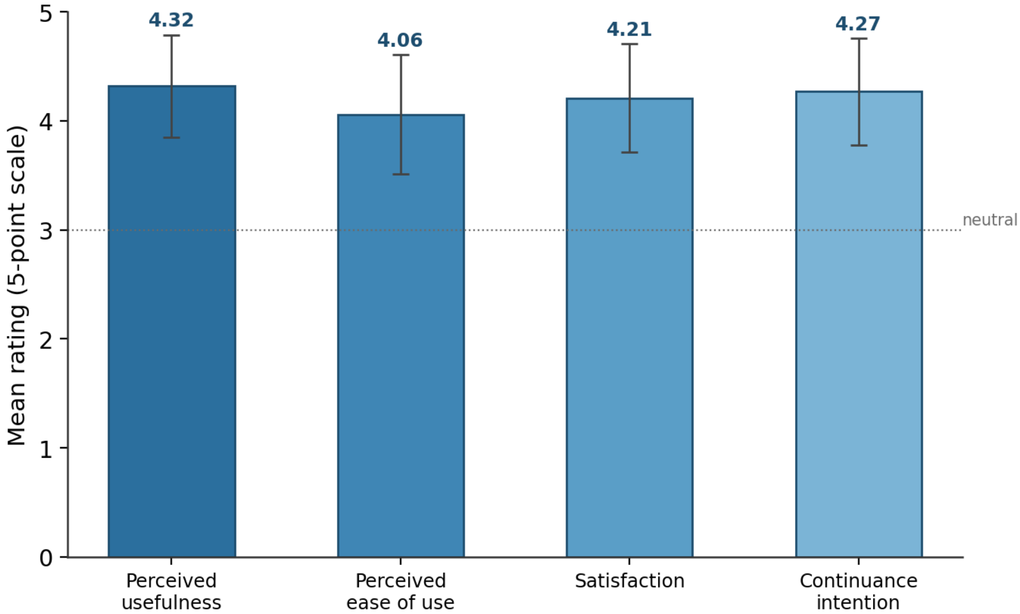

Figure 4. Technology-acceptance ratings (experimental group; bars show mean, whiskers ±1 SD).

Interview themes

Four themes emerged from the interviews. The first, which we labelled lowered barrier, captured the relief teachers felt at being handed a working example rather than a blank canvas; as one put it, starting from a case that already ran removed the fear of the empty editor. The second, scenario authenticity matters, was raised spontaneously by most interviewees: The cases landed precisely because they used real notices, real syllabi, and real datasets, and teachers contrasted this with earlier seminars whose toy examples they had found patronising. The third theme, from tools to workflows, described the conceptual shift the study set out to produce. Several teachers reported that they had stopped asking which app to open and started asking how to break a task into steps, which is as concise a statement of workflow fluency as we could have written ourselves. The fourth theme, the need for ongoing support, was the dissenting note: The gains felt fragile without a community to maintain them, and teachers worried that without continued office hours and a shared case library their new habits would lapse. This concern aligns with the slightly lower ease-of-use rating and shapes our recommendations [7,8].

Discussion

Read together, the strands tell a coherent story. A layered, case-based pathway produced competency gains far beyond those of a conventional workshop, and those gains were not merely a matter of confidence: They showed up in the time it took teachers to do real work. The largest improvements clustered where the design placed its emphasis, on building and combining workflows, which supports our central claim that the unit of faculty AI training should be the workflow rather than the tool [9]. The interviews suggest why the design worked. Exemplary cases lowered the intimidating cost of starting, authentic scenarios made the relevance self-evident, and the cumulative experience of assembling small pipelines shifted how teachers framed their problems.

These findings extend the AI-literacy literature in a practical direction. Where many studies document attitudes or self-reported awareness, our design targeted and measured transferable capability, and the efficiency data offer a behavioral anchor that attitude scales alone cannot provide. The result also speaks to the personalization gap: routing teachers to cases by diagnosed level appears to have prevented the twin failures of boring the able and overwhelming the novice, though our design cannot isolate the contribution of routing from that of the cases themselves [10].

There are clear implications for practice. Institutions that wish to move beyond one-off seminars need not buy an expensive platform; a low-code, partly open stack assembled around well-documented cases is enough, and its transparency is an asset because teachers adapt what they can see. The persistent ease-of-use concern and the interview theme of fragile gains both point to the same lesson: a training event is not a substitute for an ongoing community of practice. A shared, curated case library and light continuing support seem to be necessary complements rather than optional extras.

Several limitations temper these conclusions. The study ran at a single institution with a modest sample, so generalisation should be cautious, and the western-China setting may differ from other contexts in resourcing and policy. Group assignment was not randomised, and although we adjusted for the pre-test, unobserved differences could remain. Competency was self-reported and therefore vulnerable to social-desirability and novelty effects, even though the efficiency measures, which are harder to inflate, point the same way. The follow-up window was one semester, which cannot speak to whether the gains endure, precisely the worry teachers themselves voiced. We also cannot disentangle the effect of layered routing from the effect of the cases, an obvious target for a future factorial design. Finally, the sustained use of generative tools raises questions of academic integrity, data privacy, and over-reliance that a competency dimension can flag but not resolve.

Conclusion

Faculty AI training fails when it stops at showing teachers a tool. This study set out to test an alternative built on exemplary cases and layered, diagnosis-driven delivery, and found that early-career teachers who followed it gained substantially more competence than peers in a conventional workshop, completed authentic tasks 55-78% faster, and reported a genuine shift from tool literacy toward workflow fluency. The intervention used an accessible, largely open technology stack and is documented as a reusable case library and handbook, which makes it a plausible template for other institutions facing the same problem. The most actionable lesson is also the simplest: give teachers a real example they can run and adapt, meet them at their level, and keep a community around the work. Future research should test the design across multiple institutions, separate the effect of routing from that of the cases, and follow teachers long enough to learn whether workflow fluency, once acquired, lasts.

Funding

This work was supported by the Teacher Education Reform and Teacher Development Research Project of Baoji University of Arts and Sciences, “Research and Development of Demonstrative AI-Agent Cases for the Training of Young University Teachers and Their Implementation Pathways” (Grant No. 26JF29).

Acknowledgments

The authors would like to show sincere thanks to those techniques who have contributed to this research.

Conflicts of Interests

The authors declare no conflict of interest.

References

[1] Nguyen, D. T., Hitendra, P., Andy, Y. (2023) Understanding teachers’ knowledge and skills via technological pedagogical and content knowledge framework. *English Language Teaching*, 16(6), 91-101.

[2] Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., Wang, F. L. (2025) Retrieval-augmented generation for educational application: a systematic survey. *Computers and Education: Artificial Intelligence*, 8, 100417.

[3] Denovan, R. F., Marsasi, E. G. (2025) Perceived ease of use, perceived usefulness and satisfaction to maximize behavioral intention with the technology acceptance model in generation Y and Z consumers. *Jurnal Pamator: Jurnal Ilmiah Universitas Trunojoyo,* 18(1), 1-36.

[4] Hewavitharana, T., Nanayakkara, S., Perera, A., Perera, P. (2021) Modifying the unified theory of acceptance and use of technology (UTAUT) model for the digital transformation of the construction industry from the user perspective. *Informatics*, 8(4), 81.

[5] Braun, V., Clarke, V. (2023) Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychology Review*, 17(4), 695-718.

[6] Braun, V., Clarke, V., Hayfield, N., Davey, L., Jenkinson, E. (2023) Doing reflexive thematic analysis. *Supporting Research in Counselling and Psychotherapy: Qualitative, Quantitative, and Mixed Methods Research*, 19-38.

[7] Toropova, A., Myrberg, E., Johansson, S. (2021) Teacher job satisfaction: the importance of school working conditions and teacher characteristics. *Educational Review*, 73(1), 71-97.

[8] Benevene, P., De Stasio, S., Fiorilli, C. (2020) Well-being of schoolteachers in their work environment. *Frontiers in Psychology*, 11, 1239.

[9] Míguez-Souto, A., Gutiérrez García, M. Á., Martín-Núñez, J. L. (2025) Exploring the use of AI to optimize the evaluation of a faculty training program. *Education Sciences*, 15(10), 1394.

[10] Polomoshnov, J., Masso, A., Lobanova, K. (2026) Comparing digital, data, and AI literacy: a narrative review. *Information Polity*, 31(1), 13-26.