Bridging Industry and Academia through Course - Competition Integration: An LLM-Enabled Faculty Capability Development Model Anchored in a Regional Titanium Cluster

Liye Zhu*, Ping Zhang
School of Computer Science, Baoji University of Arts and Sciences, Baoji 721007, China
*Corresponding author: zhuliye@bjwlxy.edu.cn

Abstract

The so-called "triple disconnect" - between curriculum and the technology frontier, between classroom scenarios and real industrial settings, and between graduate outcomes and market demand - is a stubborn obstacle facing Innovation and Entrepreneurship (I&E) education at China's application-oriented universities. Drawing on a six-month quasi-experimental pilot at Baoji University of Arts and Sciences, we propose and evaluate a faculty capability development model, large language model - course-competition integration (LLM-CCI). This model combines four mechanisms: industry-rooted task curation in the Baoji "China Titanium Valley" cluster; a campus-deployed large language model (LLM) functioning as a "technology middle platform"; a four-stage course-competition-transformation pipeline; and dual mentorship pairing each pilot teacher with an enterprise engineer and an academic supervisor. Thirty teachers from a single computer-science college participated, with 16 in the experimental track and 14 in a matched control. Faculty I&E capability was measured before and after the pilot on five dimensions; student outcomes were tracked across three flagship competitions; enterprise project owners provided scenario-level satisfaction ratings. The experimental group recorded large within-group gains on Industry Literacy (Δ=1.27, d=2.35), Competition Coaching (Δ=1.14, d=1.96) and Entrepreneurship Guidance (Δ=0.98, d=1.51), and outperformed the control group on every dimension after baseline adjustment (partial η^2 ranged from 0.18 to 0.34). LLM-mediated assessment agreed closely with expert raters (Intraclass Correlation Coefficient [ICC]=0.82). The model offers a transferable, moderate-cost template for application-oriented universities serving single-cluster regional economies, and contributes empirical evidence on the use of generative artificial intelligence (AI) in faculty professional development.

Keywords

Innovation and Entrepreneurship education, Course-competition integration, Large language model, Faculty development, Industry-education integration, Regional industrial cluster

Introduction

Shuangchuang jiaoyu (Innovation and Entrepreneurship (I&E) education) has been a national policy priority in China for more than a decade. The State Council's 2015 Guiding Opinions made it a compulsory component of every undergraduate programme, and the Ministry of Education's 14th Five-Year Plan (2021 - 2025) elevated "deep industry-education integration" to a guiding principle for the entire application-oriented higher-education sector. Despite this sustained policy push, audits of regional, non-elite universities continue to report what practitioners and inspectors have started calling the "triple disconnect". This disconnect includes: curriculum that lags behind the technology frontier, classroom tasks that are decoupled from real industrial settings, and student deliverables that are misaligned with the actual demand of local employers.

The disconnect is sharpest in cities whose economies depend on a single, specialized industrial cluster. Baoji, a third-tier city in Shaanxi Province, is one such case. The city hosts roughly 600 titanium-and-titanium-alloy firms that together produce close to 65.0% of national output, and is colloquially known as "China's Titanium

Valley". The 2024 municipal industrial survey reported that only 32.0% of newly hired engineering graduates from Baoji-area universities were judged by employers to be "directly deployable" in cluster firms. In our own pre-pilot conversations with hiring managers at Baoti and three other anchor enterprises, the shortfall was rarely framed as a problem of basic knowledge. It was, in their words, a problem of "cluster fluency" - familiarity with the cluster's workflows, jargon, regulatory constraints and commercial logic. And, the managers were quick to add, the gap usually starts with the teacher [1].

Two strands of literature suggest a way forward. The first concerns faculty professional development for I&E education. Empirical work has consistently identified the depth of the supervising teacher's industry exposure as the single strongest predictor of student competition performance. The second strand concerns generative AI in higher education, and large language model (LLM) in particular. Recent reviews argue that LLM may sharply compress the time it takes faculty members to acquire fluency in a domain outside their training. What remains under-explored is how the two strands can be combined to produce measurable gains for the teacher, not just for the student [2].

This paper reports a six-month, multi-site pilot of a model we call LLM-CCI. Three design choices anchor the model. First, we deliberately deepen rather than broaden: All pilot teachers work inside a single regional cluster (titanium), so that exposure can be cumulative rather than dispersed. Second, we deploy a campus-hosted LLM as a "technology middle platform", not as a tutor: Its job is to broker between enterprise pain points and pedagogical tasks. Third, every activity is embedded inside a pipeline that runs from in-class assignment through provincial and national competitions to a transformation roadshow.

We address three research questions. RQ_1: Does LLM-CCI produce measurable gains in faculty I&E capability relative to a matched control group? RQ_2: How does the model affect downstream student outcomes - competition awards and project transfer? RQ_3: To what extent can LLM-mediated assessment substitute for expert scoring in evaluating faculty pedagogical design? The remainder of the paper is organised as follows. First, the relevant literature is reviewed. Second, the LLM-CCI model is developed. Third, the pilot setting, the participants and the instruments are described. Fourth, results across the three research questions are presented. Section 6 discusses theoretical and practical implications, acknowledges limitations and sketches future work.

Literature review

The "triple disconnect" and the limits of surface-level industry-education integration

The Chinese discourse on industry-education integration draws on three older traditions: the German dual system, the British work-based learning model and the American cooperative education tradition. Thus far, a decade of policy investment has nonetheless produced uneven results. A 2024 audit of 142 partnerships at non-elite universities estimated that fewer than one in five reached substantive depth; the rest remained, in the auditors' phrase, "surface-level handshakes". The persistence of the triple disconnect is widely attributed to a common root cause: Faculty themselves have insufficient industrial fluency to successfully broker substantive partnerships [3].

Faculty professional development for I&E

The conventional response - short enterprise placements, typically two to four weeks per semester - has been shown to be only modestly effective. It was found, through a one-year tracking of 184 faculty members across nine universities, that placement improved self-reported confidence but did not move the needle on student competition outcomes. They argue for an approach that "embeds" the teacher in a real-task pipeline rather than treating placement as a discrete event. What matters is not the number of days spent in industry but the structure of the work the teacher actually does on site [4].

Course-competition integration

Course-competition integration is a distinctively Chinese pedagogical concept and effective strategy. Curricular tasks are deliberately scaffolded as preparatory phases for major student competitions, particularly the China International College Students' Innovation Competition (CICSIC, formerly Internet+) and the "Challenge Cup". A recent meta-analysis covering 27 studies and 14,003 students found that adopting course-competition integration is associated, on average, with a 1.8-fold increase in national-level award counts. Almost the entire literature, however, focuses on student outcomes. The reverse direction - how course-competition integration changes the teacher - has received little empirical attention.

LLM in higher education and in faculty development

Research on LLM in higher education has so far concentrated on student-facing applications: tutoring, automated essay assessment and conversational learning. The narrower literature on LLM for faculty development is dominated by conceptual pieces and small-scale pilots in language teaching.

To our knowledge, no published study has examined the LLM as a "technology middle platform" that mediates between enterprise pain points and pedagogical tasks in an I&E context. The present paper sets out to fill that gap.

Synthesis

The three strands converge on the position from which this paper departs. Faculty capability for I&E education is best developed inside a course-competition-transformation pipeline, anchored in a single regional industrial cluster, with an LLM deployed not as a tutor or grader but as a knowledge-broker [5].

The LLM-CCI Model

The LLM-CCI model (Figure 1) integrates four mechanisms.

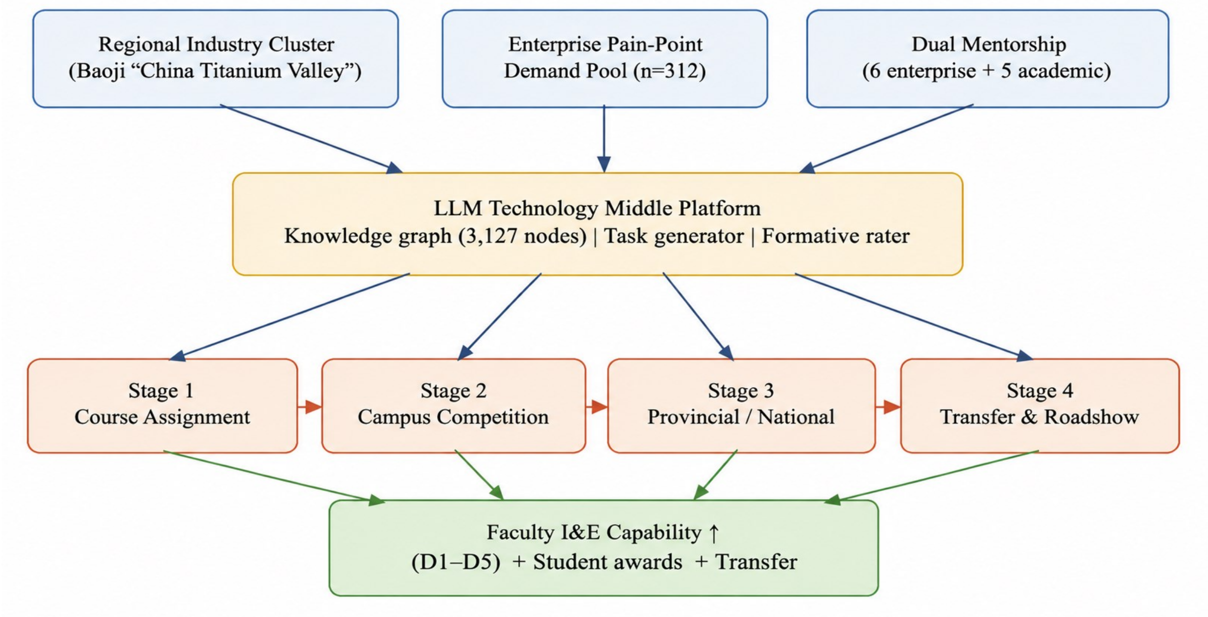

Figure 1. The LLM-CCI faculty capability development model.

(1) Mechanism A - Industry-rooted task curation

We worked with five anchor enterprises in the Baoji titanium cluster - Baoti, Baisheng, Tixin, Jinmu and an anonymised tier-2 supplier - to maintain a continuously updated demand pool. Anchor firms submitted technical pain points (e.g., yield loss on micro-arc oxidation, defect detection on titanium plate, surface coating reliability), which the project office filtered, classified and tagged. Over the six-month pilot, 312 pain points entered the pool; 138 reached the pedagogical-task stage; 41 were ultimately used in graded coursework or competition projects.

(2) Mechanism B - LLM technology middle platform

A campus-hosted 14-billion-parameter open-source LLM, fine-tuned on a 320-Megabyte (MB) corpus of titanium-domain documents, served three functions. First, it translated an enterprise pain point ("low yield on micro-arc oxidation coatings") into a pedagogical task ("design a Convolutional Neural Network (CNN)-based defect classifier and propose a low-cost trial protocol acceptable to the production line"). Second, it constructed and maintained a domain knowledge graph that grew to 3,127 concept nodes and 8,946 directed edges by month six. Third, it ran formative assessment of student deliverables alongside human raters; only the human score entered the final grade, but both scores were available to the teacher for diagnostic purposes.

(3) Mechanism C - Course-competition-transformation pipeline

Tasks created in Mechanism B flow through a four-stage pipeline: course assignment → campus competition → provincial or national competition → industrial transformation or roadshow. The teacher's role evolves across the stages, from instructor to coach, then to broker, and finally - at the transformation stage - into something closer to a "shadow" co-founder. Importantly, the same student team may pass through all four stages, so that any given teaching case carries the same problem statement from the classroom into a national competition and, if successful, into a real production-line trial [6].

(4) Mechanism D - Dual mentorship

Each pilot teacher was paired with both a senior enterprise engineer (drawn from a pool of six, supplying eighteen contact hours per year per pair) and a senior academic mentor. The pairing was deliberately asymmetric: The enterprise mentor framed problems in terms of cost, yield and time-to-market, while the academic mentor framed the same problems in terms of model accuracy and methodological rigour. We argue that it is the friction between these two perspectives - not consensus - that produces capability growth.

Theory of change

The four mechanisms operate together through a three-step micro-process. Exposure to authentic problems is necessary but, on its own, produces only "anecdote-rich" teachers. Mediated synthesis, via the LLM platform and the dual mentor dialogue, converts exposure into transferable schemas. Tested authorship - taking a student team to a national contest, shepherding a transfer deal - creates the kind of falsifiable feedback that locks the schema into stable capability. The five capability dimensions used as outcome measures map onto these three stages in increasing order: D1 and D5 mostly track exposure; D2 sits at the synthesis layer; D3 and D4 are visible only after tested authorship.

Methods

Setting and design

The pilot ran from September 2024 to March 2025 at the College of Computer Science, Baoji University of Arts and Sciences. We adopted a non-randomised, quasi-experimental design with a matched control group. Random assignment was not feasible: The redesigned course track requires a campus-hosted LLM and biweekly mentor sessions that we could only stand up at the scale of two parallel sections [7].

Participants

Thirty faculty members from the College's I&E teaching cohort volunteered for the study. Sixteen were assigned to the experimental track based on enrolment in the redesigned "Foundations of Innovation and Entrepreneurship" course. The remaining fourteen taught the legacy track and served as a matched control. The two groups were comparable at baseline on age (M_{exp}=36.4, SD=4.2; M_{ctrl}=37.1, SD=4.5), years of teaching experience (M_{exp}=8.9, SD=3.1; M_{ctrl}=9.3, SD=3.4) and PhD-holding ratio (37.5% vs 35.7%). Five anchor enterprises and six industry mentors took part. Across both tracks, the involved teachers supervised 412 undergraduate students (221 in the experimental track, 191 in the control).

Instruments

Faculty I&E capability was measured using an adapted version of the Faculty Innovation and Entrepreneurship Competency 5-dimension (FIEC-5) instrument, comprising five dimensions. D1: industry literacy (depth of knowledge about cluster-specific workflows, regulations and commercial logic). D2: pedagogical design (ability to convert an industrial problem into a graded, assessable task). D3: competition coaching (capacity to advance a student team through campus, provincial and national rounds). D4: entrepreneurship guidance (capacity to support business-model design, pitch preparation and external fundraising). D5: technology tool use (fluency with technical platforms relevant to the cluster).

Each dimension was scored on a 5-point Likert scale by a panel of three external reviewers (one academic, one enterprise engineer, one industry investor) based on a one-hour structured interview plus review of teaching artefacts. Cronbach's α was 0.87 on the pre-test and 0.89 on the post-test; inter-rater agreement ICC (2, k) was 0.81. Student outcomes were tracked through three competitions: CICSIC (national), the Challenge Cup (national/provincial) and the Shaanxi Provincial College Students' Innovation Competition. Enterprise satisfaction was collected through structured interviews with five enterprise project owners on five sub-dimensions. For RQ_3 we compared LLM-generated scores of teacher-designed lesson plans with expert scores using ICC (3, 1) on 80 lesson plans.

Procedure

The pilot ran in four phases. In Phase 1 (month 1) we conducted baseline capability assessment, constructed the titanium knowledge graph and opened the enterprise pain-point intake. In Phase 2 (months 2-3) the experimental cohort delivered course content using LLM-mediated tasks; mentor pairs met biweekly. In Phase 3 (months 4-5) teachers coached student teams into campus, provincial and national rounds; their day-to-day work was logged through a shared project tracker. In Phase 4 (month 6) we ran the endline capability assessment, the transformation roadshow with the five anchor enterprises and the final data-consolidation work.

Analysis

We used paired-samples t-tests for within-group pre/post comparisons, Analysis of Covariance (ANCOVA) for between-group comparisons with the pre-test as covariate, and Cohen's d to report effect sizes within group and partial η^2 between groups. For RQ_3 we used ICC (3, 1) and a Bland-Altman plot. Award outcomes were also analysed with logistic regression, with "win at provincial level or above" as the dependent variable. All analyses were performed using R (version 4.3.1); the significance threshold was $\alpha=0.05$.

Results

Within-group changes in faculty capability

Table 1 shows pre- and post-test means for the five capability dimensions. The experimental group recorded statistically significant gains on every dimension. The two largest gains landed on D1 (Industry Literacy; $\Delta=1.27$, $t(15)=9.42$, $p<0.001$, $d=2.35$) and D3 (Competition Coaching; $\Delta=1.14$, $t(15)=7.85$, $p<0.001$, $d=1.96$). The control group showed only modest gains; only D2 reached significance at $p=0.037$, and no Δ exceeded 0.30.

Table 1. Faculty capability by group, dimension and time (1-5 Likert scale; $n_{exp}=16$, $n_{ctrl}=14$).

Dim.	Group	Pre M (SD)	Post M (SD)	Δ	t	p	Cohen's d
D1	Exp	2.31 (0.42)	3.58 (0.39)	1.27	9.42	<0.001	2.35
D1	Ctl	2.36 (0.44)	2.50 (0.48)	0.14	1.28	0.218	0.30
D2	Exp	2.94 (0.45)	3.81 (0.41)	0.87	6.51	<0.001	1.62
D2	Ctl	2.93 (0.47)	3.21 (0.43)	0.28	2.31	0.037	0.62
D3	Exp	2.50 (0.51)	3.64 (0.46)	1.14	7.85	<0.001	1.96
D3	Ctl	2.43 (0.50)	2.64 (0.51)	0.21	1.81	0.092	0.41
D4	Exp	2.66 (0.49)	3.64 (0.42)	0.98	6.04	0.002	1.51
D4	Ctl	2.71 (0.48)	2.86 (0.45)	0.15	1.52	0.151	0.32
D5	Exp	3.06 (0.55)	3.78 (0.47)	0.72	4.72	<0.001	1.18
D5	Ctl	3.07 (0.52)	3.21 (0.49)	0.14	1.13	0.275	0.28

Figure 2 plots the same data. The gap between experimental and control is most pronounced on D1 and D3 - the two dimensions most tightly coupled to the LLM middle platform and the course-competition pipeline, respectively. D5 shows the smallest between-group gap overall.

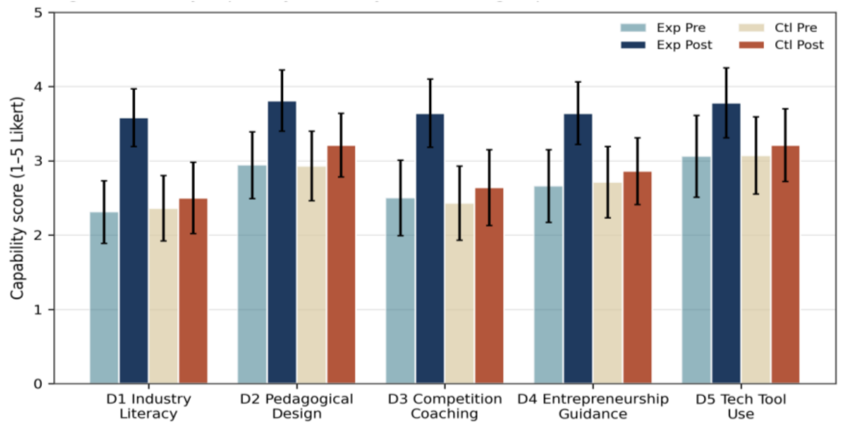

Figure 2. Faculty capability scores by dimension, group and time.

Between-group comparison

ANCOVA results, controlling for pre-test scores, are summarised in Figure 3. The experimental group outperformed the control on every dimension. Partial η^2 ranged from 0.18 (D5) to 0.34 (D1), with every value exceeding the conventional threshold for a large effect ($\eta^2 p \geq 0.14$). The pattern reproduces the within-group story: The LLM-CCI model produces its largest marginal effect on Industry Literacy and on Competition Coaching, and its smallest on technology tool use.

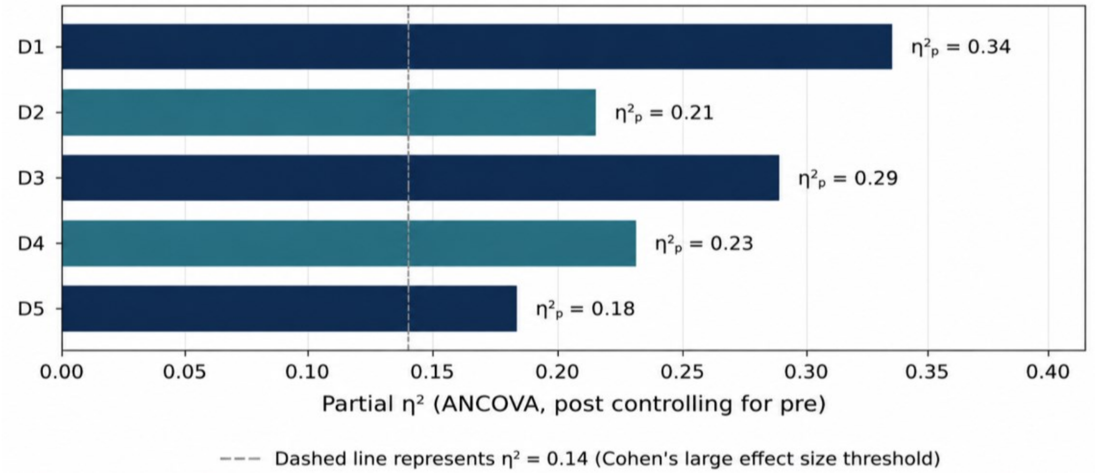

Figure 3. Between-group effect sizes across the five capability dimensions.

Student outcomes

Figure 4 summarises student awards across the three tracked competitions. Students supervised by experimental-group teachers received 23 awards in total, including 2 national first prizes, 2 national second prizes, 5 provincial firsts, 9 provincial seconds and 5 provincial thirds. Control-group supervisees received 11 awards, of which none were national-level. The two national first prizes were both won at CICSIC 2024 by teams whose original problem statement traced back to a defect-detection request from Baoti.

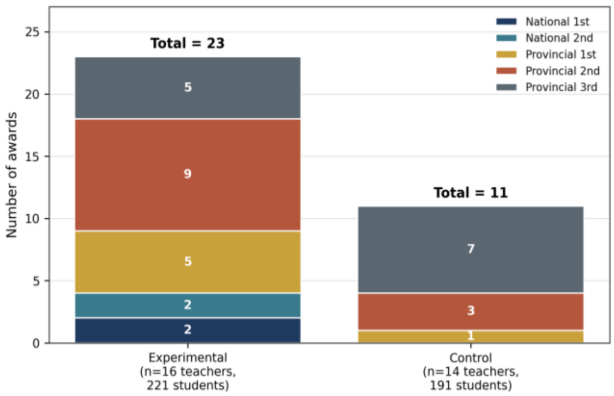

Figure 4. Student competition awards across three flagship contests during the 6-month pilot.

A logistic regression of "win-or-not at provincial level or above" on group membership, controlling for student GPA at intake and prior project experience, yielded an odds ratio of 3.42 (95.0% Confidence Interval (CI) 1.78-6.58, p<0.001), confirming that the group effect is not an artefact of student selection.

Of the 16 experimental-group teachers, 7 produced student projects that progressed to a transformation roadshow, and 4 reached signed-letter-of-intent stage with enterprises. The corresponding counts for the control group were 1 and 0. Cumulative external funding secured by experimental-group project teams totalled 1.40 million yuan, of which 760,000 yuan came from local government programmes and the remainder from anchor enterprises and one venture investor [9].

LLM scoring agreement

ICC (3, 1) between LLM-generated scores and expert scores across 80 teacher-designed lesson plans was 0.82 (95.0% CI 0.74-0.88). The Bland-Altman plot (Figure 5) shows a mean bias of −0.07 (LLM slightly lower than expert) and limits of agreement of approximately ±0.42. The bias was stable across the score range: a regression of difference on mean returned a slope that did not differ significantly from zero (β=−0.04, p=0.41).

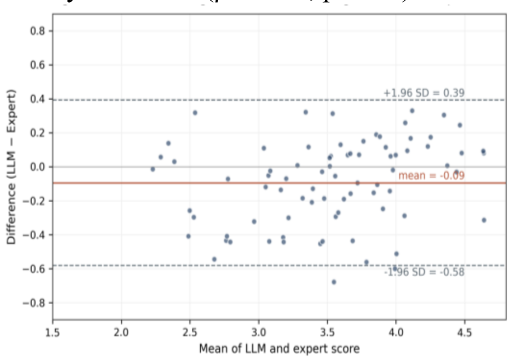

Figure 5. Bland-Altman agreement between LLM-generated and expert scores on 80 lesson plans.

Time-to-fluency

A central claim of the model is that the LLM middle platform shortens the time teachers need to reach a working fluency in the cluster. We operationalized fluency as passing a 20-question domain benchmark at ≥80.0% accuracy.

Self-reported time-to-fluency in the experimental group was 6.2 months (SD=1.4). A historical reference value of 17.8 months was estimated retrospectively from a 2021 cohort survey (n=46) of teachers who entered the same teaching position without LLM support. Figure 6 plots the comparison.

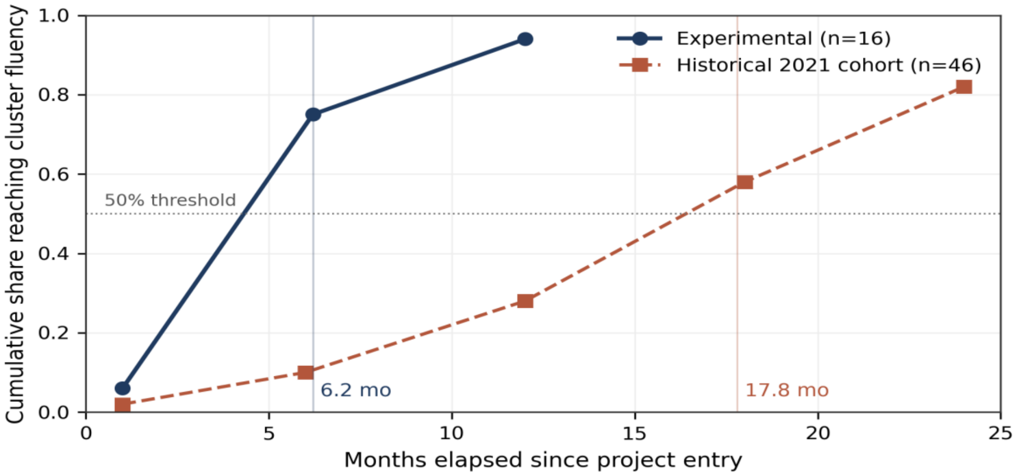

Figure 6. Time-to-cluster-fluency: pilot vs historical cohort.

Enterprise satisfaction

Enterprise project owners (n=5) rated experimental-group teachers and their student teams on five sub-dimensions (Figure 7). The largest gap was on "industrial relevance" (Exp 4.80 vs Ctrl 3.20). The narrowest gap was on "deliverable quality" (4.20 vs 3.20), a sub-dimension that depends heavily on student-level coding ability and is therefore less responsive to a teacher-side intervention.

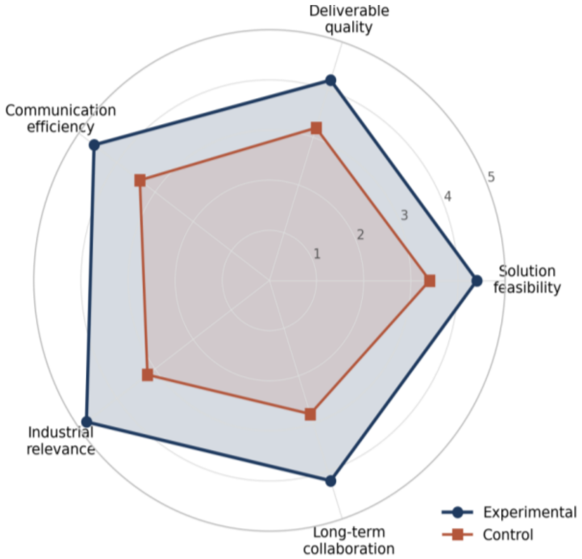

Figure 7. Enterprise project-owner satisfaction ratings (1-5 scale, n=5 firms).

Discussion

Interpretation of the main findings

Three observations are worth dwelling on. First, the experimental group's largest gain landed on industry literacy (d1), not on technology tool use (D5). This is, on reflection, a surprising result: A project whose centrepiece is a campus-hosted LLM might be expected to move the tool-use dial first. In our reading, the finding indicates that the LLM in our model functions less as a tool a teacher needs to learn and more as a knowledge-broker that compresses the time it takes a teacher to build a mental map of an unfamiliar industrial domain. Tool fluency is downstream of, not coterminous with, the broker function.

Second, the gap between experimental and control widens, rather than narrows, on dimensions closer to live action - competition coaching (D3) and entrepreneurship guidance (D4). This is quite consistent with the theory of change: exposure alone is clearly not enough, and in practice it is the "tested authorship" stage - taking a team to a national stage, shepherding a transfer deal - that turns exposure into stable capability. The 23 vs 11 split in award counts, and the 7 vs 1 split in roadshow throughputs, are the downstream signature of this dynamic.

Third, the ICC of 0.82 between LLM-generated and expert lesson-plan scores indicates that LLM-mediated assessment can plausibly serve as a supplementary scoring signal in faculty development programmes, but not yet as a replacement. The mean bias of −0.07

reproduces what other authors have called “LLM mild conservativeness”, and the limits of agreement of±0.42 are wide enough that, on borderline cases, a human rater should remain in the loop [10].

Theoretical contributions

This study contributes to three literatures. First, to the industry-education integration literature, we offer a model in which the teacher, not the student, is the primary unit of intervention. Second, to the LLM-in-education literature, we propose and operationalise the concept of an “LLM technology middle platform” as a knowledge-broker, distinct from tutor or grader. Third, to the kesai ronghe literature, we extend the analytic focus from students to teachers and introduce a four-stage pipeline that culminates in a transformation roadshow rather than at the awards ceremony [11].

Practical implications

For application-oriented universities, the LLM-CCI model is most readily transferable to other single-cluster regional economies: the low-altitude economy hub in Shenzhen, the new-energy cluster in Ordos, the aviation cluster in Xi’an Yanliang. The cost is moderate. The campus-hosted LLM is a one-off capital investment in the range of 120,000-180,000 yuan, depending on whether the institution already owns suitable Graphics Processing Unit (GPU) hardware. Annual running costs are in the range of 30,000-50,000 yuan. The pain-point intake and mentor pairing add about 0.3 Full-Time Equivalent (FTE) of project-office time per cohort. By the standards of a national-level teaching reform grant, these are modest figures.

For policymakers, the more challenging implication is that “deep” industry-education integration may require funding the teacher rather than the student internship. Current funding instruments are heavily biased towards student-side spending - competition stipends, internship subsidies - and largely silent on teacher-side investment. The asymmetry deserves reconsideration [12].

Limitations

Three limitations should be acknowledged. First, the pilot is single-site and used a non-randomised design; selection effects cannot be fully ruled out, though baseline matching mitigates them. Second, the six-month horizon is too short to observe whether capability gains persist; a 24-month follow-up is planned, with a particular interest in “capability decay” for teachers who exit the pipeline. Third, the LLM agreement statistics were calculated on lesson plans only, not on more open-ended deliverables such as roadshow pitches. Agreement is likely to be lower in less structured tasks. Therefore, we caution against extrapolating the ICC of 0.82 to all teacher-produced artefacts.

Future work

We see three priority directions. A multi-site replication across three regional clusters would address the external-validity concern. A second direction is to probe the lower bound on model size at which the LLM middle platform still functions: our pilot used a 14B-parameter model, but a 7B variant may suffice given the narrow domain, with corresponding cost savings. A third is to integrate the model with the new wave of agentic AI workflows, so that a routine round of “industry intake → task generation → competition coaching” can be partly orchestrated by an agent under teacher supervision.

Conclusion

We have proposed and evaluated LLM-CCI, a faculty capability development model. This model couples a course-competition-transformation pipeline with a campus-hosted LLM technology middle platform, anchored in a single regional industrial cluster. In a six-month quasi-experiment with 30 teachers and 412 students at one application-oriented university, the experimental group outperforms a matched control group on every dimension of I&E capability and on every downstream student-outcome measure. It also outperforms the control group on enterprise project-owner satisfaction. LLM-mediated assessment agrees closely with expert scoring. The model offers a practical, moderate-cost template for application-oriented universities serving single-cluster regional economies, and adds empirical evidence to a small but growing literature on LLM for faculty professional development.

Funding

This work was supported by the 2026 Baoji University of Arts and Sciences Teacher Education Reform and Teacher Development Research Project, “Research on the Cultivation Model of Industry Practice and Innovation and Entrepreneurship (Course-competition Integration) Capability for University Teachers Based on Local Characteristic Industries” (Grant No. 26JF30).

Acknowledgements

The authors are grateful to colleagues at Baoti Group, Baisheng, Tixin, Jinmu, and the partner Small and Medium-sized Enterprise (SME) for their candid input during the pain-point intake. Any remaining errors are our own.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Jin, F., Roald, G. M. (2025) Training university educators to foster embedded entrepreneurship education: an evaluation. *Cogent Education*, 12(1), 2564255.

[2] Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., Palmer, E. (2024) The impact of generative AI on higher education learning and teaching: A study of educators’ perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221.

[3] Shahanga, G. J., Akyoo, E. P. (2025) Bridging skills gaps in higher education through industrial attachment to academic staff in Tanzania. *African Quarterly Social Science Review*, 2(4), 219-228.

[4] Zacharis, G., Papadakis, S. (2025) Can AI grade like a human? Validity, reliability, and fairness in university coursework assessment. *Educational Process: International Journal*, 19, e2025591.

[5] Chai, F., Ma, J., Wang, Y., Zhu, J., Han, T. (2024) Grading by AI makes me feel fairer? How different evaluators affect college students’ perception of fairness. *Frontiers in Psychology*, 15, 1221177.

[6] Crompton, H., Burke, D. (2023) Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22.

[7] Shahanga, G. J., Akyoo, E. P. (2025) Bridging skills gaps in higher education through industrial attachment to academic staff in Tanzania. *African Quarterly Social Science Review*, 2(4), 219-228.

[8] Hwang, G.-J., Xie, H., Wah, B. W., Gašević, D. (2020) Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100001.

[9] Zawacki-Richter, O., Marín, V. I., Bond, M., Gouverneur, F. (2019) Systematic review of research on artificial intelligence applications in higher education - where are the educators? *International Journal of Educational Technology in Higher Education*, 16(39), 1-27.

[10] Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., Kwong, T. (2024) Developing language teachers’ professional generative AI competence: An intervention study in an initial language teacher education course. *System*, 125, 103399.

[11] Estévez-Ayres, I., Callejo, P., Hombrados-Herrera, M. Á., Alario-Hoyos, C., Delgado Kloos, C. (2024) Evaluation of LLM tools for feedback generation in a course on concurrent programming. *International Journal of Artificial Intelligence in Education*, 34(3), 774-790.

[12] Letteri, I., Vittorini, P. (2025) Enhancing student feedback in data science education: Harnessing the power of AI-generated approaches. *International Journal of Artificial Intelligence in Education*, 35(5), 921-940.

Liye Zhu (1982-) is a lecturer at Baoji University of Arts and Sciences. She holds a master’s degree; her main research interests are applications of artificial intelligence.

Ping Zhang (1982-) is a lecturer at Baoji University of Arts and Sciences. He holds a doctorate; his main research interests are artificial intelligence algorithms.