Curriculum-Embedded Customized AI Support and Learning Outcomes in University Ideological and Political Theory Courses: Evidence from an 8-Week Quasi-Experiment

Fang Zhang[1, a], Yaxin Xiao[2, b], Letong Gao[2, c], Li Yan[2, d], Yong Liang[1, e *]

[1] Department of Education, Beijing City University, Beijing, 100039, China
[2] School of Educational Science, Chengdu Normal University, Chengdu 611130, China

[a]417271414@qq.com, [b]1509756249@qq.com, [c]1103721238@qq.com, [d]2707573734@qq.com, [*e]78464980@qq.com

Abstract. Generative artificial intelligence is increasingly used in higher education, but its value may depend on whether it is embedded in course-specific pedagogy. This study examined three instructional support modes in a university ideological and political theory course. We initially recruited 163 students from three intact parallel classes; after listwise deletion, the final analytic sample included 150 students assigned to traditional teaching, traditional teaching plus general-purpose AI support, or traditional teaching plus a curriculum-embedded customized AI teaching assistant. After pretest scores were controlled, the customized AI condition showed significantly higher knowledge mastery, learning engagement, and course value recognition than both comparison conditions. The findings suggest that GenAI is more educationally useful when anchored in course resources and structured learning tasks than when used as an unrestricted general-purpose tool.

Keywords: generative artificial intelligence, customized AI teaching assistant, ideological and political theory course, learning engagement, course value recognition

1 Introduction

Generative artificial intelligence (GenAI), especially large language models (LLMs), is reshaping university feedback and after-class learning support [1-3]. Yet access to an AI tool does not by itself guarantee meaningful learning; educational effects depend on how the tool is embedded in pedagogy.

Prior studies and meta-analyses report positive effects of AI chatbots and ChatGPT-based interventions on learning performance and motivation [4, 5]. At the same time, reviews caution that poorly constrained AI use may produce inaccurate answers, shallow engagement, or overreliance [3, 6, 7].

Z. Ge et al. (eds.), *Proceedings of the 2026 6th International Conference on Education, Information Management and Service Science (EIMSS 2026)*, Atlantis Highlights in Computer Sciences 25,
https://doi.org/10.2991/978-94-6239-766-8_15

This issue is salient in ideological and political theory courses, where students must connect theoretical knowledge with interpretation, reflection, and course-level value recognition. Open-ended AI use may help with information search, but it may not align with these instructional aims.

This study compares traditional teaching, traditional teaching plus general-purpose AI support, and traditional teaching plus a curriculum-embedded customized AI teaching assistant. It asks whether course-anchored AI support is associated with stronger knowledge mastery, learning engagement, and course value recognition than either comparison condition.

2 Theoretical Background and Hypotheses

2.1 AI Support and Pedagogical Integration

GenAI can answer questions, synthesize information, and provide feedback, but these affordances become instructional only when they are connected to task goals and course materials. Without such alignment, students may treat AI as a shortcut rather than as support for deeper learning.

In this study, the customized AI teaching assistant was part of an instructional system, not a substitute for teaching. The Zhidao Teacher Version/Wisdom Tree platform connected course resources, AI-assisted interaction, and teacher-designed tasks so that students' AI use remained tied to the syllabus and weekly learning themes.

Learning theory suggests that such embedded support may improve engagement by providing structure, relevance, and feedback [8]. It may also support course value recognition by helping students relate abstract concepts to real issues and personal meaning [9].

2.2 Curriculum-Embedded Customization in Ideological and Political Theory Courses

Here, customized AI teaching assistant means curricular and pedagogical customization rather than retraining the underlying model. Customization was realized through course-resource organization, teacher-designed interaction tasks, and guided multi-turn reflection.

Learning outcomes were defined at three levels: knowledge mastery, learning engagement, and course value recognition. Together, these outcomes capture students' grasp of course knowledge, participation and cognitive investment, and recognition of the course's theoretical and practical value.

2.3 Hypotheses

General-purpose AI may support information retrieval, but its lack of course-resource constraints and dedicated pedagogical guidance may limit its contribution. A customized AI assistant embedded in an intelligent teaching platform may better align support with course objectives.

H1. Students in the traditional teaching plus customized AI teaching assistant group will demonstrate higher knowledge mastery than students in the traditional teaching group and the traditional teaching plus general-purpose AI support group.

H2. Students in the traditional teaching plus customized AI teaching assistant group will demonstrate higher learning engagement than students in the traditional teaching group and the traditional teaching plus general-purpose AI support group.

H3. Students in the traditional teaching plus customized AI teaching assistant group will demonstrate higher course value recognition than students in the traditional teaching group and the traditional teaching plus general-purpose AI support group.

3 METHOD

3.1 Participants

The study initially recruited 163 students from three intact parallel classes in the same public ideological and political theory course. During the 8-week intervention, 13 records were lost because of course withdrawal, missed posttests, or extensive missing responses on the course value recognition questionnaire. After listwise deletion, the final complete-case sample included 150 students: Group A received traditional instruction ($n = 50$; 23 males, 27 females), Group B received traditional instruction plus general-purpose AI support ($n = 51$; 24 males, 27 females), and Group C received traditional instruction plus customized AI teaching assistant support ($n = 49$; 23 males, 26 females).

In the final sample, the mean age was 19.20 years ($SD = 0.80$), and the attrition rate was about 8%. Baseline comparisons showed no significant group differences in gender distribution, chi-square = 0.013, $p = .994$, pretest scores, or prior GPA, with all p values $> .05$.

3.2 Research Design and Intervention Conditions

The study used a one-factor, three-level pretest-posttest quasi-experimental design. The independent variable was instructional support mode, with three conditions: traditional teaching, traditional teaching plus general-purpose AI support, and traditional teaching plus customized ideological and political AI teaching assistant support. The intervention lasted 8 weeks. Because intact classes were assigned to conditions, the design should be understood as a cluster-assigned classroom intervention rather than a fully randomized individual-level experiment.

Control Group A: Traditional Teaching.

Students in the control group received conventional classroom instruction. The instructor followed a unified teaching schedule and assigned regular after-class homework and reading tasks. No additional AI-based learning support was provided.

Experimental Group B: Traditional Teaching Plus General-Purpose AI Support.

In addition to regular classroom instruction, students in Group B were allowed and encouraged to use publicly available general-purpose LLM tools during after-class study for information collection, concept clarification, and discussion tasks. The teacher set basic use requirements: students were expected to complete AI-assisted learning activities around each week's theme, but no restrictions were imposed on the model's knowledge source, answer style, or interaction pathway.

The AI tools used in this condition were not connected to course-specific resources and were not embedded in a curriculum-based guidance mechanism. Their role therefore remained closer to general information support and question answering than to structured course learning.

Experimental Group C: Traditional Teaching Plus Customized AI Teaching Assistant Support.

In addition to regular instruction, Group C used a customized AI teaching assistant delivered through the Zhidao Teacher Version/Wisdom Tree platform. The assistant followed a closed-book retrieval-augmented generation (RAG) design: teacher-uploaded textbooks, lecture slides, handouts, policy documents, and case materials served as the authorized course context and sole factual anchor. Its system role was a heuristic, Socratic pedagogical agent. It was constrained not to provide final answers, but to ask students to state an initial view, provide scaffolded feedback from retrieved course materials, and pose a real-world follow-up question. If a query exceeded the uploaded course context, the assistant was expected to state this boundary and redirect the discussion to the current weekly theme. This rule operationalized the system prompt, context injection, and prompt-constraint control used throughout the intervention. Specifically, the operational prompt constraints consisted of role definition, context-only generation, and elicitation-scaffolding-deepening interaction.

The platform recorded task completion and interaction participation for instructional monitoring, but it did not provide standardized participant-level logs that could be matched across all students. These process data were therefore not analyzed as outcome variables.

3.3 Measurement Instruments

Knowledge Mastery Test.

Knowledge mastery was assessed with a standardized course test developed by the teaching and research office. The test was based on the course syllabus, instructional objectives, and core chapter content. It covered key concepts, foundational theories, and the application of theory to practice. The test was scored on a 100-point scale and included objective items and material-analysis questions. Objective items assessed understanding and retention of fundamental course knowledge, whereas material-analysis questions assessed students' ability to analyze real-world issues using course theories.

A pretest and a posttest were administered before the intervention and at the end of Week 8, respectively. The two tests were designed to be broadly equivalent in content coverage, item structure, and difficulty. Material-analysis questions were scored with a standardized rubric to reduce subjective scoring error as far as possible.

Learning Engagement.

Learning engagement was measured using the Utrecht Work Engagement Scale-Student (UWES-S) [10]. The scale contains 17 items across three dimensions: vigor, dedication, and absorption. Responses were collected on a 5-point Likert scale ranging from 1 (very inconsistent) to 5 (very consistent), with higher scores indicating greater learning engagement. Participants were instructed to respond based on their recent experience in this course to improve contextual fit.

In the present sample, the overall Cronbach's alpha was .89, indicating good internal consistency. Confirmatory factor analysis supported the expected three-factor structure, with satisfactory model fit: chi-square/df = 1.742, CFI = .984, TLI = .981, RMSEA = .046, and SRMR = .027.

Course Value Recognition.

Course value recognition (CVR) was defined as the degree to which students understand, endorse, and accept the theoretical significance, practical value, and personal developmental relevance of ideological and political theory courses. This construct differs from general course satisfaction and broad value-attitude measures because it focuses on students' internal understanding and recognition of the course content and its practical significance during learning.

The questionnaire was developed with reference to measurement approaches used in political psychology, course identification, and ideological and political education research. Items were generated according to the instructional goals and syllabus of the course, then reviewed and revised by ideological and political theory teachers and educational researchers for clarity, course relevance, and dimensional appropriateness. The final questionnaire contained 15 items across three dimensions: cognitive recognition, affective recognition, and behavioral tendency. A 5-point Likert scale was used, with higher scores indicating stronger recognition of course value. The final three-factor model showed satisfactory fit: chi-square/df = 1.869, CFI = .944, TLI = .931, RMSEA = .036, and SRMR = .021.

Cognitive recognition reflects students' understanding of the course's theoretical significance and practical function; affective recognition reflects acceptance of and emotional orientation toward the course content; and behavioral tendency reflects willingness to think about and analyze real-world issues using perspectives learned in the course. In the present study, the scale yielded a Cronbach's alpha of .91, indicating good internal consistency.

3.4 Procedure

The intervention lasted 8 weeks. In Week 1, all participants completed baseline measures, including the knowledge mastery test, the learning engagement scale, and the course value recognition questionnaire. From Weeks 2 to 7, the intervention was implemented under the three instructional support conditions. Across all groups, classroom content, teacher assignment, and teaching progress were kept consistent; differences lay primarily in after-class support. In Week 8, all participants completed a unified posttest, and the relevant questionnaire and course assessment data were collected.

3.5 Data Analysis

Records were screened before analysis. The raw pool contained 163 students; cases with withdrawal, missed posttests, or extensive missing questionnaire data were excluded through listwise deletion/complete case analysis, and no imputation was used. The final sample was 150. Data were analyzed in SPSS 27.0 using descriptive statistics, reliability analysis, and one-way ANCOVA, with posttest outcomes as dependent variables, corresponding pretest scores as covariates, and instructional support mode as the independent variable. Significant ANCOVA results were followed by Bonferroni comparisons.

4 Results

4.1 Differences in Knowledge Mastery

To test Hypothesis 1, a one-way ANCOVA was conducted with posttest knowledge mastery as the dependent variable, pretest knowledge mastery as the covariate, and instructional support mode as the independent variable. After pretest performance was controlled, instructional support mode was significantly associated with knowledge mastery, $F(2, 146) = 15.34$, $p < .001$, partial eta-squared = .17, indicating a moderately large group difference.

As shown in Table 1, the customized AI teaching assistant group (Group C) had an adjusted mean of 88.45 (SE = 0.65), compared with 83.15 (SE = 0.64) for the general-purpose AI support group (Group B) and 81.65 (SE = 0.65) for the traditional teaching group (Group A). Bonferroni post hoc comparisons showed that Group C scored significantly higher than both Group B and Group A (both p values < .001). The difference between Groups A and B was comparatively small. Figure 1(a) visualizes the group pattern. These findings are consistent with Hypothesis 1 and suggest that curriculum-embedded customized AI support was associated with stronger knowledge mastery than either comparison condition.

Table 1. Descriptive Statistics and ANCOVA Results for Learning Outcomes by Instructional Support Mode.

Variable	Condition	Pretest M (SD)	Posttest M (SD)	Adjusted M (SE)	F(2, 146)	p	Partial eta-squared	Post Hoc (Bonferroni)
KM	Group A	71.20 (5.30)	81.60 (5.40)	81.65 (0.65)	15.34	< .001	.17	C > A***, C > B***
	Group B	70.80 (5.50)	83.20 (5.10)	83.15 (0.64)				

Variable	Condition	Pretest M (SD)	Posttest M (SD)	Adjusted M (SE)	F(2, 146)	p	Partial eta-squared	Post Hoc (Bonferroni)
	Group C	71.50 (5.10)	88.50 (4.20)	88.45 (0.65)				
LE	Group A	3.25 (0.42)	3.65 (0.52)	3.66 (0.07)	11.28	< .001	.13	C > A**, C > B*
	Group B	3.28 (0.45)	3.78 (0.61)	3.77 (0.07)				
	Group C	3.22 (0.40)	4.12 (0.45)	4.13 (0.07)				
CVR	Group A	3.30 (0.48)	3.75 (0.51)	3.76 (0.06)	22.67	< .001	.23	C > A***, C > B***
	Group B	3.32 (0.45)	3.82 (0.47)	3.81 (0.06)				
	Group C	3.28 (0.46)	4.35 (0.38)	4.36 (0.06)				

[a] KM = knowledge mastery; LE = learning engagement; CVR = course value recognition. Group A = traditional teaching (n = 50); Group B = traditional teaching + general-purpose AI (n = 51); Group C = traditional teaching + customized AI teaching assistant (n = 49). Pretest scores were entered as covariates in the ANCOVA models. Adjusted M represents the estimated marginal mean after controlling for pretest scores. * p < .05, ** p < .01, *** p < .001.

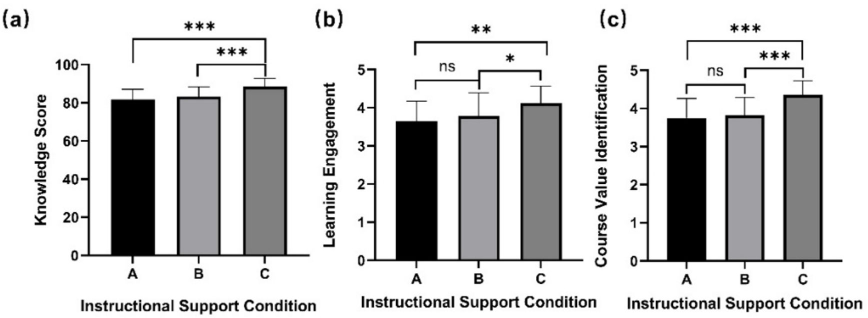

Fig. 1. Differences in learning outcomes across instructional support conditions: (a) posttest knowledge mastery, (b) learning engagement, and (c) course value recognition.
Note. Error bars represent standard deviations. Group A = traditional teaching; Group B = traditional teaching + general-purpose AI; Group C = traditional teaching + customized AI teaching assistant. ns = non-significant; * p < .05; ** p < .01; *** p < .001.

4.2 Differences in Learning Engagement

To test Hypothesis 2, a one-way ANCOVA was conducted with posttest learning engagement as the dependent variable, pretest learning engagement as the covariate, and instructional support mode as the independent variable. Instructional support mode was significantly associated with posttest learning engagement, $F(2, 146) = 11.28$, $p < .001$, partial eta-squared = .13, indicating a moderate group difference.

Table 1 shows an adjusted mean of 4.13 (SE = 0.07) for Group C, compared with 3.77 (SE = 0.07) for Group B and 3.66 (SE = 0.07) for Group A. Bonferroni comparisons indicated that Group C scored significantly higher than Group A ($p < .01$) and Group B ($p < .05$). By contrast, Groups A and B did not differ significantly ($p = .34$). Figure 1(b) presents the group pattern. These findings are consistent with Hypothesis 2 and indicate that higher engagement was observed only when AI support was embedded in a structured, course-based, and guided learning process.

4.3 Differences in Course Value Recognition

To test Hypothesis 3, a one-way ANCOVA was conducted with posttest course value recognition as the dependent variable, pretest course value recognition as the covariate, and instructional support mode as the independent variable. After pretest levels were controlled, instructional support mode was significantly associated with course value recognition, $F(2, 146) = 22.67$, $p < .001$, partial eta-squared = .23, indicating a relatively strong group difference.

Table 1 shows an adjusted mean of 4.36 (SE = 0.06) for Group C, whereas the adjusted means for Group B and Group A were 3.81 (SE = 0.06) and 3.76 (SE = 0.06), respectively. Bonferroni post hoc comparisons showed that Group C scored significantly higher than both Group B and Group A (both p values < .001). The difference between Groups A and B was relatively small. Figure 1(c) illustrates the group pattern. These results are consistent with Hypothesis 3 and suggest that curriculum-embedded customized AI support was linked to stronger course value recognition.

5 Discussion

This study compared traditional teaching, general-purpose AI support, and curriculum-embedded customized AI teaching assistant support. After baseline levels were controlled, the customized AI condition showed the strongest outcomes across knowledge mastery, learning engagement, and course value recognition. The results suggest that AI value depends less on access alone than on alignment with course resources, tasks, and instructional goals [3, 6, 7].

5.1 Interpretation of Findings

For knowledge mastery, curriculum-embedded customized AI support was associated with higher adjusted posttest scores. Course-anchored explanations and prompts may have helped students focus on key concepts rather than use AI mainly as a shortcut. For learning engagement, the customized condition also showed a clearer advantage: students had to articulate initial ideas, interact around weekly themes, and refine their understanding. For course value recognition, the result should be read as course-level recognition, not broad value-system change. The assistant may have helped students connect abstract concepts with real-world issues and personal meaning.

5.2 Theoretical Contributions

The study contributes by distinguishing general-purpose AI access from curriculum-embedded AI support, extending AI-in-education research to ideological and political theory courses, and showing that responsible GenAI use requires alignment among instructional aims, resources, and student tasks.

5.3 Practical Implications

Practically, instructors should not treat GenAI only as a question-answering tool. Existing intelligent teaching platforms can be used to connect course materials, task requirements, and guided interaction so that AI support reduces superficial use and better serves course objectives.

5.4 Limitations and Future Research

Several limitations should be noted. First, intact-class assignment limits causal inference. Second, the single-university, single-course sample restricts generalizability. Third, course value recognition was measured with a study-developed scale; although reliability and model fit were acceptable, its stability should be further tested across samples.

Fourth, knowledge mastery was measured with a formal course test, but learning engagement and course value recognition relied on self-report scales and may be affected by social desirability or subjective response drift. Because standardized LMS logs such as login frequency, time-on-task, forum posting, and resource-access records were unavailable for all participants, future studies should use these indicators to triangulate self-reported outcomes.

Fifth, the customized assistant operated within the Zhidao/Wisdom Tree infrastructure, whose foundation-model backbone parameters were not fully disclosed. Therefore, we cannot directly benchmark its semantic reasoning capacity against GPT-4, Claude, or similar commercial APIs. The findings should be read as evidence for a curriculum-embedded, knowledge-anchored support design, not for a specific model's superiority. Future work should compare backbone models, retrieval strategies, and prompt constraints under equivalent conditions.

6 Conclusion

General-purpose AI support did not automatically produce stronger engagement or course value recognition. Curriculum-embedded customized AI support, however, was associated with stronger knowledge mastery, engagement, and value recognition. GenAI appears most useful when aligned with course objectives, resource constraints, task design, and guided interaction.

References

1. O. Zawacki-Richter, V. I. Marin, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education: Where are the educators?" International Journal of Educational Technology in Higher Education, vol. 16, no. 1, article 39, 2019, doi: 10.1186/s41239-019-0171-0.
2. H. Crompton and D. Burke, "Artificial intelligence in higher education: The state of the field," International Journal of Educational Technology in Higher Education, vol. 20, no. 1, article 22, 2023, doi: 10.1186/s41239-023-00392-8.
3. E. Kasneci, K. Sessler, S. Kuchemann, M. Bannert, D. Dementieva, F. Fischer, et al., "ChatGPT for good? On opportunities and challenges of large language models for education," Learning and Individual Differences, vol. 103, article 102274, 2023, doi: 10.1016/j.lindif.2023.102274.
4. R. Wu and Z. Yu, "Do AI chatbots improve students' learning outcomes? Evidence from a meta-analysis," British Journal of Educational Technology, vol. 55, no. 1, pp. 10-33, 2024, doi: 10.1111/bjet.13334.
5. R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," Computers & Education, vol. 227, article 105224, 2025, doi: 10.1016/j.compedu.2024.105224.
6. Y. Qian, "Pedagogical applications of generative AI in higher education: A systematic review of the field," TechTrends, vol. 69, pp. 1105-1120, 2025, doi: 10.1007/s11528-025-01100-1.
7. F. Miao and W. Holmes, Guidance for Generative AI in Education and Research. Paris: UNESCO, 2023, doi: 10.54675/EWZM9535.
8. E. R. Kahu, "Framing student engagement in higher education," Studies in Higher Education, vol. 38, no. 5, pp. 758-773, 2013, doi: 10.1080/03075079.2011.598505.
9. M. Vansteenkiste, N. Aelterman, G.-J. De Muynck, L. Haerens, E. Patall, and J. Reeve, "Fostering personal meaning and self-relevance: A self-determination theory perspective on internalization," The Journal of Experimental Education, vol. 86, no. 1, pp. 30-49, 2018, doi: 10.1080/00220973.2017.1381067.
10. W. B. Schaufeli, I. M. Martinez, A. M. Pinto, M. Salanova, and A. B. Bakker, "Burnout and engagement in university students: A cross-national study," Journal of Cross-Cultural Psychology, vol. 33, no. 5, pp. 464-481, 2002, doi: 10.1177/0022022102033005003.