Constructing an AI-Assisted Teaching Model for Data Analysis and Data Visualization Courses in Higher Education: An Empirical Study Based on Student Performance Data

Ziwen Wang[a*], Chunxu Li[b], Huisi Li[c]

School of Cyberspace Security, Guangzhou University of Software, Guangzhou, China
[a*]bitter_mango@163.com, [b]18749289591@163.com
[c]sandyli93@163.com

Abstract. The integration of AI tools in higher education is accelerating, yet few studies systematically embed them into data analysis and visualization instruction. This paper proposes the AI-DAV teaching model, a three-dimensional, AI-TPACK-grounded closed-loop framework spanning AI-assisted teaching, learning, and assessment. Deployed over a semester in an ND2009 course at a Chinese university (five classes, 244 students), the model was evaluated via a practicum project scored on five dimensions: data preprocessing, analysis, visualization, technique application and description, and report writing. The evaluation results reveal that Students scored highest in data visualization (80.4%), which received the most direct AI support, and lowest in data preprocessing (75.6%). One-way ANOVA showed significant inter-class differences ($F(4,239) = 2.84$, $p = 0.025$, $\eta^2 = 0.045$), with the Information Systems Development class significantly outperforming others. The between-class gap was narrowest in visualization (5.7 points) and widest in preprocessing (8.9 points)—consistent with a compensatory effect, although bootstrap contrasts are not statistically distinguishable at this sample size. These findings empirically validate the model and yield actionable implications for AI-era instructional design.

Keywords: AI-assisted teaching model; data analysis and data visualization; AI-TPACK; higher education; educational data mining; instructional framework

1 Introduction

The proliferation of large language models (LLMs) and AI programming tools is reshaping the teaching of programming and data science courses in higher education. Since the release of ChatGPT in late 2022, a growing body of research has documented the impacts of generative AI on student learning outcomes. A meta-analysis of 35 experimental studies (4193 participants) reported a moderate positive overall effect of ChatGPT on learning outcomes ($g = 0.670$, 95% CI [0.495, 0.844]) [1]. In an evalua-

© The Author(s) 2026
M. B. Badzis et al. (eds.), *Proceedings of the 2026 2nd International Conference on Education Research and Training Technologies (ERTT 2026)*, Advances in Social Science, Education and Humanities Research 1042,
https://doi.org/10.2991/978-2-38476-628-4_7

tion on past CS1 exam questions, OpenAI Codex solved roughly 80% of the problems, placing it within the top quartile of the enrolled cohort [2]. In the field of data visualization, Chen et al. [3] systematically evaluated GPT's capabilities for Harvard's CS171 course. Their findings revealed that the model can not only generate code, but also clean and explore CSV datasets, read SVG-format visualization outputs, and produce explanatory visualizations to articulate data insights.

Despite these advances, most empirical studies target CS1/CS2 courses or general pedagogy; few explore systematic AI integration into integrated courses combining data analysis and visualization. Such courses demand two distinct competencies—statistical thinking (cleaning, transformation, analysis) and visual design (chart construction, encoding, narrative)—which rely on different cognitive processes and hence receive differentiated AI support. Moreover, frameworks such as AI-TPACK [5] and Intelligent-TPACK [4] remain largely conceptual, lacking empirical validation in discipline-specific contexts.

This paper aims to address this gap by proposing the AI-DAV (AI-assisted Data Analysis and Visualization) teaching model, a three-dimensional framework that enables systematic integration of AI tools across the complete instructional cycle: teaching, learning, and assessment. The model was deployed over a full semester within a "Data Analysis and Data Visualization" course at a Chinese university, involving five classes with 244 students in total. Multidimensional analysis of student practicum project performance data was conducted to empirically validate the proposed model.

This study addresses three core research questions:

- RQ1 (Model Design): What are the constituent components, structural relationships, and operational mechanisms of an AI-assisted teaching model tailored for data analysis and data visualization courses?
- RQ2 (Implementation Effects): After implementing the model in real teaching scenarios, what patterns can be observed in student performance across different competency dimensions?
- RQ3 (Model Reflection): Drawing on student performance data, which dimensions of the model deliver the strongest effects, and which require iterative optimization?

This paper makes three contributions. First, it proposes the domain-specific AI-DAV model, operationalizing AI-TPACK into actionable instructional design for data analysis and visualization courses. Second, it provides classroom evidence from five classes (244 students) based on multidimensional project assessment rather than subjective self-reports. Third, it formalises a reliance diagnostic and an equity-gap estimation procedure, identifying differentiated AI support across competency dimensions, with results suggesting that AI can narrow achievement gaps in specific competencies.

2 Related Work

2.1 AI Tools in Programming and Data Education

AI code generators have shown considerable proficiency on CS1-level tasks [2], prompting educators to rethink assignment design. GPTutor, a ChatGPT-powered tool, offers in-IDE code explanation [6]; its authors identify the relationship between usage frequency and course performance as future work. In data visualization, Chen et al. [3] showed that GPT can clean CSV datasets, generate code for multiple libraries, interpret SVG outputs, and construct explanatory visualizations—extending AI assistance beyond code generation into core visualization pedagogy. At the tool level, Microsoft Research's Data Formulator combines GUI interaction with natural-language input [7]. Collectively, these tools now support the full data-analysis workflow, from cleaning and visual exploration to narrative construction.

2.2 AI-Assisted Teaching Frameworks

The Technological Pedagogical Content Knowledge (TPACK) framework [8] has long grounded educational technology research. Ning et al. [5] proposed the AI-TPACK framework, decomposing teacher knowledge into components including AI Technological Knowledge (AI-TK), AI Technological Content Knowledge (AI-TCK), and AI Technological Pedagogical Knowledge (AI-TPK). In a complementary line, Celik [4] proposed Intelligent-TPACK (I-TPACK), which extends TPACK with teachers' knowledge for the ethical assessment of AI-based decisions. From an implementation perspective, AI-U [9] builds on LLM fine-tuning and retrieval-augmented generation to adapt to instructors' styles, and GAIDE [10] supports course-material development. A shared consensus is that AI should act as a collaborator rather than a replacement. Nevertheless, most frameworks target general pedagogy or CS1/CS2, and few are tailored to the integrated data analysis and visualization course.

2.3 Learning Analytics and Educational Data Mining

In assessment and analytics, Saez et al. [11] proposed a framework and case study for AI-assisted assignment grading in higher education using large language models, establishing an empirical basis for AI-assisted grading. Educational data mining and learning analytics have been widely applied to predict student performance [12], and gradient-boosting models such as XGBoost have proven effective for this task [13]. A systematic literature review of learning-analytics dashboards [14] identified persistent limitations: weak alignment with pedagogical theory, neglect of teachers' usage contexts, and overemphasis on technical function. Analytical modules should therefore be embedded in coherent pedagogical frameworks rather than deployed as isolated tools. Yet few studies validate domain-specific AI-assisted models for data-analysis and visualization courses using multidimensional project performance data.

2.4 Research Gaps

Synthesizing the above literature review, four prominent research gaps are identified within existing scholarship:

(1) Course misalignment: Most studies target CS1/CS2 courses; few address AI-assisted teaching for the integrated data-analysis and visualization course despite its distinct cognitive demands.

(2) Theory–validation gap: Frameworks such as AI-TPACK and I-TPACK emphasize theory and teacher competency; limited work operationalizes them into actionable models validated with student performance data.

(3) Narrow assessment: LLM-assisted grading proves technical viability but lacks a closed loop spanning AI-assisted teaching, learning, and assessment.

(4) Weak feedback: Learning analytics enable prediction but seldom feed insights back into iteratively refining AI-assisted teaching models.

This study addresses all four gaps simultaneously. It develops a domain-specific teaching model for the integrated course of data analysis and data visualization, which is empirically validated using multidimensional performance data collected from five classes comprising 244 students.

3 The AI-DAV Teaching Model

3.1 Theoretical Foundation

The AI-DAV model is grounded in the AI-TPACK framework [5], which integrates AI-specific knowledge dimensions into the original TPACK model [8]. Within the context of data analysis and data visualization integrated courses, the three core dimensions of AI-TPACK are instantiated as follows:

- **AI-TK**: Instructors' awareness of AI tool functions and limitations, and operational proficiency with AI programming assistants (Tongyi Lingma), LLM visualization advisors (Tongyi Qianwen), and the intelligent teaching platform (Zhitu Haoke).
- **AI-TCK**: Matching AI capabilities to course content—e.g., assistants excel at NumPy and Matplotlib generation, while LLM advisors support exploratory analysis and visualization design decisions.
- **AI-TPK**: Pedagogical strategies, e.g., assignments requiring critical evaluation of AI-generated code, scaffolding from guided to independent AI use, and platform analytics to identify at-risk students.

Drawing on the above AI-TPACK theoretical basis, this study constructs the three-dimensional closed-loop AI-DAV teaching model covering AI-assisted teaching, AI-assisted learning, and AI-assisted assessment.

3.2 Model Architecture

The AI-DAV model consists of three interrelated dimensions that jointly form a closed-loop instructional structure: AI-Assisted Teaching (AI-AT), AI-Assisted Learning (AI-AL), and AI-Assisted Assessment (AI-AA). Figure 1 demonstrates the overall architectural framework of the proposed model, while Table 1 elaborates on the core functions and supporting tools corresponding to each dimension.

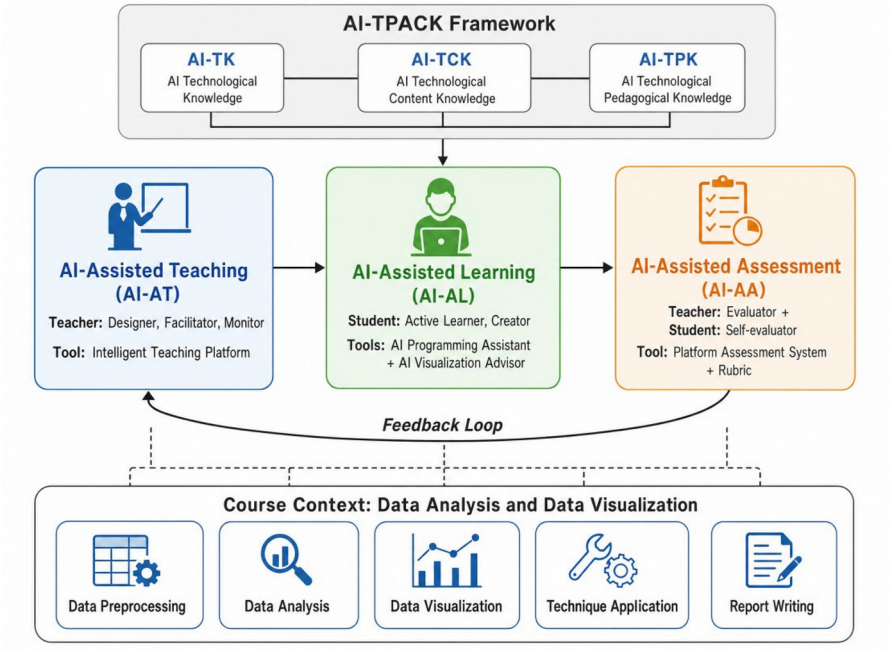


Fig. 1. The AI-DAV teaching model framework, illustrating its three-dimensional closed-loop structure derived from the AI-TPACK framework and aligned with the five assessment dimensions defined for this course.

AI-Assisted Teaching (AI-AT). Instructors use the intelligent teaching platform Zhitu Haoke to deliver content, distribute assignments, and monitor learning. The platform records video viewing, quiz completion, and resource access, and offers visualized dashboards that help instructors identify at-risk students, shifting their role from knowledge transmitter to instructional designer and facilitator.

AI-Assisted Learning (AI-AL). Students use two categories of AI tools in laboratories and projects: AI programming assistants (Tongyi Lingma), an IDE-embedded code completion tool supporting Pandas, NumPy, Matplotlib, Seaborn, and Pyecharts tasks; and the LLM-based conversational agent Tongyi Qianwen, a visualization advisor that assists with chart selection, visual encoding, pattern interpretation, and analytical narratives. Students are required to critically examine, verify, and optimize AI outputs rather than adopt them uncritically, cultivating critical thinking.

AI-Assisted Assessment (AI-AA). A five-dimensional rubric scores practicum projects: data preprocessing (15), data analysis (25), data visualization (25), analytical technique application and description (15), and report writing (20). Assessment combines platform-captured process data with instructor rubric scores, enabling fine-grained analysis of differentiated AI support across competency dimensions.

Table 1. Structure of the AI-DAV Teaching Model

Dimension	Teacher Role	Student Role	AI Tools	Key Activities
AI-Assisted Teaching (AI-AT)	Designer, facilitator, monitor	—	Intelligent teaching platform (Zhitu Haoke)	Content delivery, progress tracking, adaptive resource recommendation
AI-Assisted Learning (AI-AL)	Guide, scaffolding provider	Active learner, tool user, creator	AI programming assistant; AI visualization advisor	Code writing, data exploration, chart design, debugging
AI-Assisted Assessment (AI-AA)	Evaluator, feedback provider	Self-evaluator	Platform assessment system, AI-assisted scoring rubric	Multidimensional scoring, progress feedback, remedial instruction

3.3 Closed-Loop Mechanism

The three dimensions of the AI-DAV model are not independent stages, but form a continuous closed-loop feedback mechanism. Assessment outcomes from AI-AA feed back into the AI-AT instructional design process, enabling instructors to optimize teaching focuses and refine AI tool integration strategies. For example, if multidimensional assessment data reveal sustained student underperformance in the data preprocessing dimension, instructors can supplement targeted instructional activities that require students to diagnose data quality problems and rationalize data cleaning strategies, rather than merely conducting mechanical code execution. This iterative teaching optimization represents a core feature of the design-based research methodology adopted in this study.

To make this closed-loop mechanism executable rather than purely descriptive, we formalise it as Algorithm 1. For each class, the algorithm computes a standardised shortfall g_{kd} for every assessment dimension, triggers interventions from a predefined instructional library $\mathcal{L}$ whenever a shortfall exceeds a threshold δ, and additionally activates a metacognitive intervention when the density of reliance-flagged students (from Algorithm 3) in a class exceeds ρ_{max}. The resulting intervention plans are ranked by priority, giving instructors an actionable, data-driven basis for the next instructional cycle.

Table 2 lists the instructional activities in the intervention library $\mathcal{L}$ used by the algorithm, aligned with the five assessment dimensions and the metacognitive safeguard recommended in Section 5.2.

Table 2. Intervention library $\mathcal{L}$ used by Algorithm 1

Trigger	Instructional activity
Preprocessing	Defect-seeded dataset diagnosis; compare and justify competing AI-generated cleaning schemes
Analysis	Method-selection rationale worksheet before coding
Visualisation	Critique of AI-proposed encodings against task type
Technique desc.	Structured oral defence of analytic choices
Report writing	Peer review against the rubric, AI output prohibited
Metacognitive	Predict-then-verify protocol: commit to an expected result before querying the AI tool

Algorithm 1 Closed-Loop Instructional Trigger(AI-AA→AI-AT)

Input: Class–dimension means $\bar{P} \in \mathbb{R}^{K\times D}$; cohort means $\bar{p}_d$, SDs σ_d; intervention library $\mathcal{L}$; thresholds δ, ρ_{max}; flags $\mathcal{F}^+$ from Alg.3

Output: Ranked intervention plan Π_k per class

1: for k←1 to K do
2: $\Pi_k \leftarrow [\,]$
3: for d←1 to D do
4: $g_{kd} \leftarrow (\bar{P}_{kd} - \bar{p}_d)/\sigma_d$
5: if $g_{kd} < -\delta$ then
6: Append($\Pi_k, \langle \mathcal{L}(d), |g_{kd}| \rangle$)
7: end if
8: end for
9: $\rho_k \leftarrow |\{i \in \mathcal{F}^+ : c_i = k\}|/n_k$
10: if $\rho_k > \rho_{max}$
11: Append($\Pi_k, \langle \mathcal{L}(\mathrm{METACOG}), \rho_k \rangle$)
12: end if
13: SortByPriorityDesc(Π_k)
14: end for
15: return $\{\Pi_k\}_{k=1}^{K}$

3.4 Course Context Adaptation

The AI-DAV model targets the "Data Analysis and Data Visualization" course (ND2009), a compulsory course for Information Management and Information Systems undergraduates. It spans 54 class hours (26 theoretical, 28 laboratory) across eight progressive modules (Python, NumPy, Pandas, Matplotlib, Seaborn, Pyecharts), culminating in a practicum project in which each student independently produces a 3,000–5,000-word analysis report. This project-driven structure trains the full compe-

tency chain—acquisition, preprocessing, analysis, visualization, communication—where AI tools deliver differentiated support across stages.

4 Methodology

4.1 Research Design

This study adopts Design-Based Research (DBR) [15], an iterative framework integrating model design, classroom implementation, data analysis, and instructional revision. This paper reports the first complete cycle: model design, semester-long deployment, and empirical analysis of practicum performance data. As a single-cycle implementation without a control group, analysis relies on within-group comparisons across five dimensions and between-group comparisons among five natural classes.

4.2 Participants and Research Context

The empirical implementation was conducted during the Spring 2026 semester (March–July) at a Chinese university, covering five student classes with differentiated academic backgrounds:

(1) Class 1 (IS Development track): 57 students (43 M, 14 F), regular undergraduates in Information Management and Information Systems.

(2) Class 2 (Outstanding Engineer / IRM tracks): 39 students (27 M, 12 F), regular undergraduates from relevant majors.

(3) Classes 3–5 (Top-up degree track): three classes of 50, 49, and 49 students with vocational backgrounds—richer practical experience but weaker theoretical foundations.

All five classes shared the same teacher, content, AI-DAV implementation, AI tools, and rubric. The valid sample was $N = 244$.

4.3 AI Tool Integration

Three categories of AI tools were embedded to support the closed-loop framework:

(1) Intelligent teaching platform (Zhitu Haoke): content delivery, resource distribution, quiz administration, and progress monitoring; records online performance (10% of grade) and provides teaching dashboards.

(2) AI programming assistant (Tongyi Lingma): an IDE-embedded code completion tool for Python preprocessing, computation, and visualization tasks.

(3) AI visualization advisor (Tongyi Qianwen): an LLM conversational tool assisting chart selection, visual encoding, pattern interpretation, and report structuring.

Students received standardized guidance emphasizing verification and optimization of AI outputs, logical understanding rather than mechanical copying, and academic integrity.

4.4 Data Collection

The core data source was the comprehensive practicum project, worth 50% of the course grade, in which students independently completed full-process analysis (acquisition, preprocessing, statistical analysis, visualization, report writing). It was scored with a standardized five-dimensional rubric consistent with the model's competency framework:

- **Data preprocessing (15)**: data cleaning, missing-value and outlier handling, standardized transformation.
- **Data analysis (25)**: rational method selection, statistical reasoning, insight extraction.
- **Data visualization (25)**: chart suitability, visual design, presentation effectiveness.
- **Analytical technique application and description (15)**: interpreting technical logic, justifying analytical decisions.
- **Data analysis report writing (20)**: report structure, narrative coherence, academic standardization.

The course grade comprises regular assignments (25%), laboratory exercises (10%), classroom performance (5%), online learning (10%), and the practicum project (50%). This study analyzes practicum data, since the multidimensional rubric enables fine-grained evaluation of differentiated AI support across competency dimensions.

4.5 Data Analysis Methods

Descriptive statistics (mean, SD, range) were computed per dimension and for the total score; grades were categorized as Excellent (90–100), Good (80–89), Fair (70–79), Pass (60–69), and Fail (<60). One-way ANOVA tested inter-class differences, with Games-Howell post-hoc comparisons (unequal variances) and effect sizes η^2 and Cohen's d; raw dimension scores were converted to percentages, and significance was set at $p < 0.05$.

Because the compensatory claim concerns relative between-class dispersion across dimensions, we estimate η^2 per dimension and obtain confidence intervals by stratified bootstrap (resampling within class to preserve the unbalanced design), formalised in Algorithm 2. The paired contrast reuses each replicate for all dimensions, so the correlation among dimensions is retained and testing Δ_{uv} directly is more appropriate than comparing marginal intervals. With $B = 10{,}000$, $n = 244$, and $D = 5$, it runs in under a second.

Algorithm 2 Bootstrap Dimension-wise Equity Gap (BDEG)

Input: Normalised scores P=[pid]; class labels ci∈{1,…,K}; replicates B; level α
Output: $\hat{\eta}^2_d$ with CI for each d; pairwise contrasts
1: for d←1 to D do
2: $\hat{\eta}^2_d \leftarrow ETASQ(p_d, c)$
3: end for
4: for b←1 to B do

$I_b \leftarrow 0$
for k←1 to K do
$I_b \leftarrow I_b \cup RESAMPLE(\{i: c_i = k\}, n_k)$
end for
for d←1 to D do
$\eta_v^{2(b)} \leftarrow$ETASQ(p_{I_b}, d, c_{I_b})
end for
end for
for d←1 to D do
$C_{I_d} \leftarrow [\, Q_{\alpha/2}(\eta_d^{2(\cdot)}),\ Q_{1-\alpha/2}(\eta_d^{2(\cdot)}) \,]$
end for
for all dimension pairs (u,v) do
$\Delta_{uv}^{(b)} \leftarrow \eta_u^{2(b)} - \eta_v^{2(b)}$ for $b = 1, \ldots, B$
$p_{uv} \leftarrow 2\min\left(\Pr[\Delta_{uv}^{(b)} \le 0], \Pr[\Delta_{uv}^{(b)} \ge 0]\right)$
end for
return $\{\hat{\eta}_d^2, CI_d\}, \{p_{uv}\}$

Algorithm 3 Dimension-ProfileRelianceDiagnostic(DPRD)

Input: Score matrix $S \in \mathbb{R}^{n\times D}$; max-point vector $m \in \mathbb{R}^D$; AI-automatability ratings $a \in [0,1]^D$; tail quantile α

Output: Reliance index $r \in \mathbb{R}^n$; flag sets $\mathcal{F}^+, \mathcal{F}^-$

Step 1: scale-free normalisation

for d←1 to D do
for i←1 to n do
$p_{id} \leftarrow S_{id}/m_d$
end for
$\mu_d \leftarrow \mathrm{mean}(p_{\cdot d}), \sigma_d \leftarrow \mathrm{sd}(p_{\cdot d})$
for i←1 to n do
$z_{id} \leftarrow (p_{id} - \mu_d)/\sigma_d$
end for
end for

Step 2: reliance index

$A \leftarrow \sum_d a_d, C \leftarrow \sum_d (1 - a_d)$
for i←1 to n do
$\bar{z}_i^{auto} \leftarrow \frac{1}{A}\sum_d a_d\, z_{id}$
$\bar{z}_i^{conc} \leftarrow \frac{1}{C}\sum_d (1 - a_d) z_{id}$
$r_i \leftarrow \bar{z}_i^{auto} - \bar{z}_i^{conc}$

Step 3: distribution-free flagging

$\tau^+ \leftarrow \mathrm{quantile}(r, 1-\alpha), \tau^- \leftarrow \mathrm{quantile}(r, \alpha)$
$\mathcal{F}^+ \leftarrow \{i: r_i \ge \tau^+\}$
$\mathcal{F}^- \leftarrow \{i: r_i \le \tau^-\}$

19: return $\mathrm{r}, \mathcal{F}^{+}, \mathcal{F}^{-}$

4.6 Reliance Diagnostic

A single aggregate score cannot distinguish a student who has mastered the analytic workflow from one who delegates artefact production to AI while never developing the underlying judgement. To operationalise the diagnosis in Section 5.2, this paper introduce the Dimension-Profile Reliance Diagnostic (DPRD) (Algorithm3). Step 1 removes the differing rubric maxima (15–25 points) so dimensions become comparable in standard-deviation units. Step 2 contrasts standardised performance on artefact-producing dimensions against judgement-oriented ones, weighted by an externally elicited AI-automatability rating a rather than the authors' assumption. Step 3 flags using empirical quantiles, since r is not assumed normal. The procedure runs in $O(nD)$ time and space and needs only existing rubric scores, so it transfers to any multidimensional-rubric course.

The vector a is designed to be elicited from J instructors, each rating every dimension on a 0–1 scale for how far current AI tools can produce an acceptable artefact with minimal student judgement. A provisional vector was used here; formal elicitation and reporting of inter-rater agreement (ICC(2,J)) is left to the next design cycle (Section 6).

4.7 Descriptive Statistics

Table 3 summarizes the descriptive statistics of overall project scores for the full sample and each individual class.

The overall average score was 78.29 (SD = 8.10, range 55–96). Class 1 achieved the highest mean (81.04) with the lowest SD (6.69), while Class 2 obtained the lowest mean (75.77) with the greatest dispersion (SD = 9.84). The three top-up degree classes scored moderately (77.46–78.18). Only two students, both in Class 4 (55 and 56 points), failed the project.

Table 3. Descriptive Statistics for Total Project Score

Group	N	Mean	SD	Min	Max	Range
Class 1 (IS Dev.)	57	81.04	6.69	70	95	25
Class 2 (OE/IRM)	39	75.77	9.84	60	96	36
Class 3 (Sec. 01)	50	77.46	8.46	60	92	32
Class 4 (Sec. 03)	49	78.04	7.75	55	89	34
Class 5 (Sec. 04)	49	78.18	7.14	65	95	30
Total	244	78.29	8.10	55	96	41

Table 4. Grade Distribution

Grade	Range	C1	C2	C3	C4	C5	Total	%
Excellent	90-100	10	3	4	0	3	20	8.2%
Good	80-89	25	9	17	22	23	96	39.3%
Fair	70-79	22	17	22	24	16	101	41.4%
Pass	60-69	0	10	7	1	7	25	10.2%
Fail	<60	0	0	0	2	0	2	0.8%

Table 5. Mean Scores by Assessment Dimension(N=244)

Dimension	Max	Mean	SD	% of Max	Rank
Data Visualization	25	20.09	3.02	80.4%	1
Data Analysis	25	19.74	2.91	79.0%	2
Technique APP.& Description	15	11.68	1.84	77.9%	3
Report Writing	20	15.44	2.31	77.2%	4
Data Preprocessing	15	11.34	1.72	75.6%	5

4.8 Grade Distribution

Table 4 illustrates student grade distribution categorized by the five-tier grading standard.

Only two students (0.8%) failed, both in Class 4; 80.7% attained Good or Fair. No Class 1 student fell into Pass or Fail, whereas 10 Class 2 students (25.6%) scored 60–69, and failure rates among the top-up classes ranged from 2.0% to 14.3%.

4.9 Dimensional Analysis

Table 5 presents the means and percentages of the maximum score for each assessment dimension across all 244 students.

Students scored highest in data visualization (80.4% of full marks), the dimension receiving the most direct AI support—assistants generate Matplotlib/Seaborn/Pyecharts code and advisors guide chart design—and lowest in data preprocessing (75.6%), which demands conceptual cognition of data quality, rational cleaning decisions, and context-dependent judgment that AI supports only indirectly. The 4.8-point gap demonstrates differentiated AI effects: output-centric dimensions benefit most, while judgement-centric ones benefit least.

4.10 Class Comparison

Table 6 presents the comparison across the five classes for each dimension, along with the inter-class range (in percentage points) for each dimension, to assess the compensatory effect.

One-way ANOVA on the total score revealed statistically significant differences among the five classes: $F(4,239) = 2.84$, $p = 0.025$, with a small effect size ($\eta^2 = 0.045$). Games-Howell post hoc tests indicated that Class 1 (Information Systems Development track) significantly outperformed Class 2 (Outstanding Engineer/IRM track) ($p < 0.01$, $d = 0.63$) and also significantly outperformed the combined group of the three top-up degree classes (Classes 3–5) ($p < 0.01$, $d = 0.43$). No significant difference was found between Class 2 and the combined top-up degree group ($p = 0.215$, $d = 0.26$).

Dimension-wise, the between-class gap was smallest in data visualization (5.7 points), where AI support is most direct, and largest in data preprocessing (8.9 points), which demands conceptual judgment—consistent with a compensatory effect. As Section 4.11 reports, however, the bootstrap contrasts are not statistically distinguishable at $K = 5$ clusters, so the ordering is suggestive rather than confirmatory.

Unexpectedly, Class 5—second-lowest in total score—ranked highest in the technique application and description dimension (80.2%), surpassing Class 1 (76.4%). This reversal suggests top-up students, with vocational practical experience, are more adept at articulating technical choices, a competency that benefits more from hands-on practice than from theory.

Table 6. Inter-Class Comparison by Dimension

Dimension	C1(%)	C2(%)	C3(%)	C4(%)	C5(%)	Range
Data Preprocessing	81.2	72.3	72.9	77.7	72.4	8.9
Data Analysis	81.8	74.9	79.9	78.5	78.4	6.9
Data Visualization	83.8	79.6	80.3	79.2	78.1	5.7
Technique APP.& Description	76.4	74	78.5	79.5	80.2	6.2
Report Writing	80.0	75.9	73.4	75.2	80.8	7.4
Total	**81.0**	**75.8**	**77.5**	**78**	**78.2**	**5.3**

Table 7. Dimension-wise between class η^2 with bootstrap 95% CI

Dimension	η^2	CI low	CI high
Data Preprocessing	0.120	0.064	0.217
Data Analysis	0.055	0.019	0.146
Data Visualization	0.043	0.017	0.118
Technique APP.& Description	0.052	0.019	0.136
Report Writing	0.101	0.054	0.191

4.11 Reliance Profile Analysis

Algorithm 2 with 10,000 stratified replicates yielded dimension-wise η^2 from 0.043 (data visualization) to 0.120 (data preprocessing); the confidence intervals overlap substantially (Table 7). The paired contrast between visualization and preprocessing gives $\Delta\eta^2 = 0.077$, $p = 0.050$. The ordering is consistent with a compensatory pat-

tern, but at $K = 5$ clusters the dimension-wise contrasts are not statistically distinguishable—a qualification that must accompany any compensatory claim.

Figure 2 visualises these estimates as a forest plot.

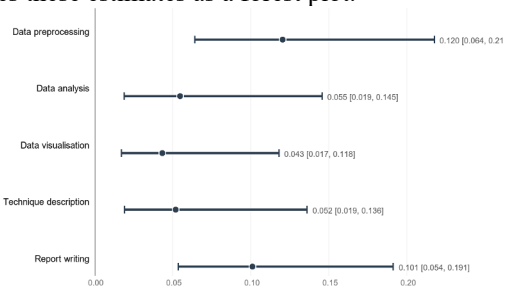


Fig. 2. Dimension-wise between-class η^2 with bootstrap 95% confidence intervals (Algorithm 2).

Table 8. Distribution of $\mathcal{F}^+$ (reliance-flagged) students by class

Class	N	$\mathcal{F}^+$	Density
Class 1 (IS Dev.)	57	6	10.5%
Class 2 (OE/IRM)	39	14	35.9%
Class 3 (Sec. 01)	50	12	24%
Class 4 (Sec. 03)	49	2	4.1%
Class 5 (Sec. 04)	49	4	8.2%

The DPRD diagnostic (Algorithm 3, $\alpha = 0.15$) produced a reliance index r with mean 0.00, SD 0.31 (range $[-0.85, 0.63]$); the distribution appears in Figure 3. Upper-tail students ($\mathcal{F}^+$, output-strong but judgement-weak, $n = 38$) and lower-tail ($\mathcal{F}^-$, judgement-strong but output-weak, $n = 37$) were flagged. Crucially, the mean total score of $\mathcal{F}^+$ (79.45) was not lower than that of the remaining students (78.20) or of $\mathcal{F}^-$ (77.51). This is the central argument for multidimensional assessment: AI-reliant students are invisible in a single aggregate score but become visible in the dimension profile, with $\mathcal{F}^+$ students scattered across the full range of total scores.

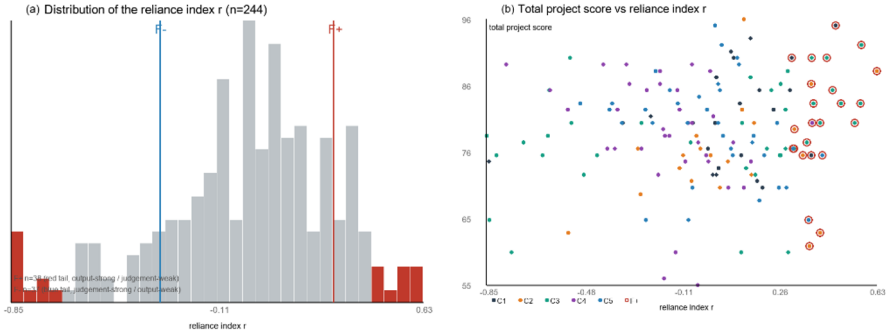


Fig. 3. (a) Distribution of the reliance index r (Algorithm 3); the red and blue tails mark $\mathcal{F}^+$ and $\mathcal{F}^-$. (b) Total project score versus reliance index r, coloured by class; $\mathcal{F}^+$ students are outlined.

The distribution of $\mathcal{F}^+$ students across classes is uneven (Table 8). Class 2 (Outstanding Engineer/IRM track) has the highest reliance-flag density (35.9%), followed by Class 3 (24.0%), whereas Class 4 (4.1%) and Class 5 (8.2%) show low densities. This pattern provides a concrete, actionable signal for the closed-loop trigger (Algorithm 1): Class 2 and Class 3 would receive metacognitive interventions in addition to any dimension-specific ones.

4.12 Summary of Findings

The empirical results respond to the three research questions as follows:

RQ1: The AI-DAV model organizes AI integration into three interconnected dimensions—AI-assisted teaching, learning, and assessment—each specifying tool arrangements and roles, forming a closed-loop structure linking curriculum content, AI capabilities, and multidimensional evaluation.

RQ2: Student performance shows a stable trend: data visualization, the most directly AI-supported dimension, scored highest (80.4%), while data preprocessing, relying on conceptual judgment, scored lowest (75.6%). This held across all five classes, with visualization ranking first or second everywhere.

RQ3: The model supports tangible output creation (visualization code, charts) most strongly but fosters conceptual judgment least. The between-class gap was smallest in visualization (5.7 points), consistent with a compensatory effect, though the bootstrap contrasts (Section 4.11) show the ordering is not statistically distinguishable at this sample size. Inter-class differences were significant yet small ($\eta^2 = 0.045$), suggesting the model provides relatively homogeneous support across diverse cohorts.

5 Discussion

5.1 Interpretation of Results

Across all five classes, students scored highest in data visualization, consistent with AI tools providing the most direct support there: assistants generate valid Matplotlib, Seaborn, and Pyecharts code, while LLM advisors recommend chart types and encodings. In contrast, the weakest dimension, data preprocessing, requires diagnosing data-quality problems and selecting cleaning strategies—domain-specific judgment that AI code generation cannot replace, and for which the current model offers insufficient scaffolding.

The visualization pattern merits attention: the between-class gap of 5.7 points there contrasts with 8.9 points in preprocessing. If confirmed with larger samples, this would indicate that AI can equalize highly structured, automatable tasks. However, as the bootstrap contrast in Section 4.11 shows, the evidence is suggestive rather than conclusive, and disparities in fundamental analytical competencies persist. The pattern still carries practical implications for educational equity and warrants confirmatory study.

Finally, top-up degree students outperforming regular undergraduates in the analytical technique application and description dimension—which demands articulating method choices and discussing limitations—suggests that metacognitive interpretation of technical reasoning benefits more from hands-on vocational practice than theoretical instruction, underscoring the multidimensional nature of data analysis competency.

5.2 Implications for the AI-DAV Model

Three targeted optimizations follow. First, AI-Assisted Learning should embed structured activities on data preprocessing—e.g., diagnosing defects in deliberately flawed datasets before adopting AI-recommended cleaning code, or comparing and justifying competing AI-generated schemes—strengthening conceptual scaffolding in the weakest dimension. Second, AI-Assisted Teaching can use platform analytics to identify students over-dependent on AI (qualified outputs but weak independent judgment on preprocessing) and arrange targeted interventions; dimension-level evaluation enables this fine-grained identification that a single aggregate score cannot. Third, AI-Assisted Assessment should add AI-enabled formative feedback during project development, not only summative scoring of final deliverables, prioritizing domains where standalone AI support is weakest.

5.3 Comparison with Existing Research

The overall trend aligns with meta-analytic evidence that generative AI exerts moderate positive effects on learning ($g = 0.670$) [1]. Notably, this pattern is not a simple restatement of prior meta-analytic findings: Wu et al. [1] report a slightly larger pooled effect for declarative knowledge ($g = 0.903$) than for procedural knowledge ($g = 0.705$), although the difference was not significant ($Q = 1.599$, $p = 0.206$). Our dimension-level results therefore describe a distinction—between artefact production and conceptual judgement—that cuts across the declarative/procedural dichotomy used in prior meta-analyses and that a single aggregate outcome measure cannot resolve. The findings also support Chen et al. [3], who showed LLMs can complete multiple phases of visualization workflows, confirming the viability of embedding AI tools in such courses.

The significant yet small inter-class effect ($\eta^2 = 0.045$) echoes evidence that student background moderates technology-enhanced learning [14], while showing the model delivers relatively stable support across groups. The visualization pattern tentatively extends prior work by suggesting AI may narrow gaps in specific competencies; because the dimension-wise contrasts were not statistically distinguishable (Section 4.11), this hypothesis requires confirmatory work with more classes.

5.4 Limitations

This study has several limitations. First, the absence of a control group restricts causal inference: without a comparable non-AI cohort, performance variations cannot be

definitively attributed to the model, as instructor effects, curriculum design, and motivation may contribute; cross-dimensional and inter-class comparisons thus offer only indirect evidence. Second, although the sample ($N = 244$) exceeds most single-course studies and covers two student populations, it comes from one university and one course, limiting generalizability. Third, only final project scores were analyzed; process data (AI usage logs, learning trajectories) were not incorporated. Fourth, no perception or self-efficacy questionnaire/interview data were collected, limiting explanations of dimensional differences, especially the score reversal in the technique application dimension. Future work can address these via longitudinal tracking, control groups, process-data collection, and mixed-method designs.

6 Conclusion and Future Work

This paper proposes the AI-DAV teaching model, a three-dimensional framework embedding AI tools across teaching, learning, and assessment for data analysis and visualization courses, deployed over one semester across five classes (244 students). Students scored highest in data visualization (most direct AI support) and lowest in data preprocessing (demanding conceptual judgment). Inter-class differences were significant ($F(4,239) = 2.84$, $p = 0.025$, $\eta^2 = 0.045$), verifying the moderating effect of educational background; the between-class gap narrowed to 5.7 points in visualization, consistent with a compensatory effect, though bootstrap contrasts show the ordering is not yet statistically distinguishable.

These findings validate the model and yield three suggestions: match AI tools purposefully to competency dimensions (prioritizing data preprocessing); supplement AI assistance with targeted pedagogical interventions for conceptual competencies; and adopt multidimensional rubrics to capture differentiated AI effects and identify over-reliant students.

Future work proceeds along three trajectories: (1) a second design-based research cycle implementing the proposed optimizations, within which the AI-automatability vector a (Algorithm [alg:dprd]) will be formally elicited from J instructors and their inter-rater agreement ICC(2,J) reported, replacing the provisional vector; (2) analysis of platform and AI-usage process data to deepen insight into student–AI interactions; and (3) controlled experiments with a non-AI-assisted comparison group to strengthen causal inference. Through iterative cycles, the model can mature into an empirically grounded, generalizable solution for AI-assisted data science education.

Acknowledgments

This work was supported by the Ministry of Education Industry-Academia Collaborative Education Program (No. 250701950034359), and the Scientific Research Projects of Guangzhou University of Software under Grant No. KY202420.

References

1. X. Wu, P. Zhu, J. Zhang, M. Yin, and Y. Wang, “ChatGPT’s impact on student learning outcomes: A meta-analysis of 35 experimental studies,” Humanities and Social Sciences Communications, vol. 13, no. 1, article 684, 2026. https://doi.org/10.1057/s41599-026-07019-z.
2. J. Finnie-Ansley, P. Denny, B. A. Becker, A. Luxton-Reilly, and J. Prather, “The robots are coming: Exploring the implications of OpenAI Codex on introductory programming,” in Proc. 24th Australasian Computing Education Conf. (ACE ’22), 2022, pp. 10–19. https://doi.org/10.1145/3511861.3511863.
3. Z. Chen, C. Zhang, Q. Wang, J. Troidl, S. Warchol, J. Beyer, N. Gehlenborg, and H. Pfister, “Beyond generating code: Evaluating GPT on a data visualization course,” in 2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis), 2023, pp. 16–21. https://doi.org/10.1109/EduVis60792.2023.00009.
4. I. Celik, “Towards Intelligent-TPACK: An empirical study on teachers’ professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education,” Computers in Human Behavior, vol. 138, article 107468, 2023. https://doi.org/10.1016/j.chb.2022.107468.
5. Y. Ning, C. Zhang, B. Xu, Y. Zhou, and T. T. Wijaya, “Teachers’ AI-TPACK: Exploring the relationship between knowledge elements,” Sustainability, vol. 16, no. 3, article 978, 2024. https://doi.org/10.3390/su16030978.
6. E. Chen, R. Huang, H.-S. Chen, Y.-H. Tseng, and L.-Y. Li, “GPTutor: A ChatGPT-powered programming tool for code explanation,” in Proc. 24th Int. Conf. Artificial Intelligence in Education (AIED 2023), LNCS, Springer, 2023, pp. 321–327. https://doi.org/10.1007/978-3-031-36336-8_50.
7. C. Wang, J. Thompson, and B. Lee, “Data Formulator: AI-powered concept-driven visualization authoring,” IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 1, pp. 1128–1138, 2024. https://doi.org/10.1109/TVCG.2023.3326585.
8. P. Mishra and M. J. Koehler, “Technological pedagogical content knowledge: A framework for teacher knowledge,” Teachers College Record, vol. 108, no. 6, pp. 1017–1054, 2006. https://doi.org/10.1111/j.1467-9620.2006.00684.x.
9. M. F. Shojaei, R. Gulati, B. A. Jasperson, S. Wang, S. Cimolato, D. Cao, W. Neiswanger, and K. Garikipati, “AI-University: An LLM-based platform for instructional alignment to scientific classrooms,” arXiv preprint arXiv:2504.08846, 2025. https://arxiv.org/abs/2504.08846.
10. E. Dickey and A. Bejarano, “GAIDE: A framework for using generative AI to assist in course content development,” in Proc. IEEE Frontiers in Education Conf. (FIE), 2024. https://doi.org/10.1109/FIE61694.2024.10893132.
11. Y. Saez, L. M. Garcia, A. Mochon, and P. Isasi, “Evaluating large language models for AI-assisted grading: A framework and case study in higher education,” Scientific Reports, vol. 16, no. 1, 2026. https://doi.org/10.1038/s41598-026-48656-3.
12. C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” WIREs Data Mining and Knowledge Discovery, vol. 10, no. 3, article e1355, 2020. https://doi.org/10.1002/widm.1355.
13. A. Asselman, M. Khaldi, and S. Aammou, “Enhancing the prediction of student performance based on the machine learning XGBoost algorithm,” Interactive Learning Environments, vol. 31, no. 6, pp. 3360–3379, 2021. https://doi.org/10.1080/10494820.2021.1928235.

14. M. Sahin and D. Ifenthaler, "Visualizations and dashboards for learning analytics: A systematic literature review," in Advances in Analytics for Learning and Teaching, Springer, 2021, pp. 3–22. https://doi.org/10.1007/978-3-030-81222-5_1.
15. A. L. Brown, "Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings," Journal of the Learning Sciences, vol. 2, no. 2, pp. 141–178, 1992. https://doi.org/10.1207/s15327809jls0202_2.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

