Journal of Economic Education and Entrepreneurship Studies
Vol. 7, No. 5, Oktober 2026, pp. 1-16
ISSN: 2722-3744 (Print) 2776-0278 (Online)
https://doi.org/10.62794/je3s.v7i5.306

Optimizing AI-Driven Feedback Through Learning Style Alignment: Evidence from a Factorial Experiment among Accounting Students

Diyah Probowulan[1*], Ilham Saifudin[2], and Nina Martiana[3]

[1,3]Department of Accounting, Universitas Muhammadiyah Jember, East Java, Indonesia
[2]Department of Informatic Engineering, Universitas Muhammadiyah Jember, East Java, Indonesia

*Corresponding Author
Jl. Karimata No 49, Jember – Jawa Timur, Indonesia, 68121
e-mail: diyahprobowulan@unmuhjember.ac.id

Received: July 10, 2026; **Revised:** August 25, 2026; **Accepted**: September 05, 2026

Abstract: Accounting students often work on computationally intensive procedural tasks such as journal entries and depreciation schedules, where feedback is often delayed and generic, so it is unclear whether AI-based feedback interacts with learners' processing preferences. This study examines how three levels of AI-based feedback, rule-based direct feedback, retrospective ML personalization feedback, and prospective ML predictive feedback combine with learning styles (visual, kinesthetic, and convergent, classified through the Kolb Learning Style Inventory 4.0) to shape accounting learning outcomes. A 3×3 factorial experiment between subjects (N = 135) assigned students to one of nine combinations of force-based feedback. Because pre-test scores differed systematically across different pre-intervention feedback conditions, learning improvements (after the test minus pre-tests), rather than post-test scores, became the primary outcome, with naïve post-test and covariate analyses reported as convergence checks, followed by Tukey's HSD comparisons in which the omnibus effect was significantly relevant. Feedback conditions, learning styles, and interactions each contributed significantly to learning improvements, but the patterns were not uniform: some learning style groups gained relative gains regardless of feedback levels, while others showed feedback-dependent improvements that did not consistently support more technologically advanced conditions. These results suggest that AI-based feedback values are dependent on learner characteristics, not universal, and simpler feedback is not necessarily inferior after baseline differences are taken into account. Given the simple single-location sample, unresolved fundamental imbalances under various conditions, and reliance on composite outcome sizes, these findings should be treated as preliminary evidence justifying replication before force-based feedback configurations are adopted in practice.

Keywords: AI-driven feedback, personalized learning, learning styles, accounting students, adaptive learning

How to Cite: Probowulan, D., Saifudin, I., & Martiana, N. (2026). Optimizing AI-Driven Feedback Through Learning Style Alignment: Evidence from a Factorial Experiment among Accounting Students. *Journal of Economic Education and Entrepreneurship Studies*, *7*(5), 1–16. https://doi.org/10.62794/je3s.v7i5.306

INTRODUCTION

Accounting education has a persistent feedback problem: students work through graded journal entries and depreciation schedules, then wait days before figuring out where their reasoning is broken, so procedural misunderstandings are already consolidated (Sundkvist & Kulset, 2024). Machine learning (ML)-based AI systems offer a structurally different approach—learning from each student's response history, detecting repeated error patterns, and adjusting instructional content in real time—placing them in a very different class from the rules-based automated feedback platforms that have been available for decades (Bhattacharya, 2025; Mellado-Silva et al., 2020).

Cognitive Load Theory (CLT) explains why feedback mechanisms should be important: working memory operates under strict capacity constraints, and adaptive or predictive ML-based feedback serves as a framework that suppresses additional burdens by providing directional guidance calibrated according to each student's current knowledge conditions (Renkl & Atkinson, 2003; Sweller et al., 2011; Wang, 2024). A second theoretical lens—Kolb's experiential learning theory, explains for whom sophisticated feedback is most effective. Visual learners construct meaning through diagrams and annotated solutions; kinesthetic learners require iteration and hands-on practice; Convergent learners integrate abstract reasoning with active application, a profile that naturally fits into accounting's dual demands of conceptual precision and procedural execution (Andayani et al., 2026; Eloff, 2017; Stanciu et al., 2020; Suriyanti et al., 2026). Importantly, affective responses to feedback are not peripheral: meta-analytical work (Mertens et al., 2022; H. Wang & Lehman, 2021) shows that the motivational effects of personalized feedback often outweigh cognitive effects.

The literature on AI feedback and learning styles in accounting runs on separate paths (Ebaid, 2022; Gunarathne et al., 2020). Their interactions, whether certain levels of ML-based feedback are effective differently depending on learning styles have not been tested in factorial designs. These gaps are consequential: (Bhattacharya, 2025) argues that mismatches between feedback mechanisms and learning orientation can dampen or reverse expected performance improvements. This factorial blend of instructional treatment (AI feedback level) and learner characteristics (Kolb's learning style) follows the broader tradition of talent-treatment interaction (ATI) in educational psychology, which argues that instructional effectiveness depends on a match between the learner's treatment and talents, rather than on one best treatment for all learners (Cronbach & Snow, 1977). The study addressed the gap through a 3×3 factorial experiment, making three contributions: (1) the first controlled comparison of the three levels of ML-based feedback in accounting education; (2) the first factorial test of learning style as a moderator of the effectiveness of AI feedback; and (3) an empirically grounded discussion of when the sophistication of feedback helps, hinders, or does not make a difference, as a basis for more careful system design.

AI-Driven Automated Feedback: Conceptual Clarification and Theoritical Grounding

It is important to distinguish the AI-based feedback examined here, which is operationalized at three increasingly personalized levels from conventional automated feedback systems such as McGraw-Hill Connect or WileyPlus, which despite providing direct truth feedback, adaptive test delivery, and remedial item targeting based on rule sets and item response data, still execute fixed programmatic rules rather than learning and updating dynamically as per the behavior of individual students (Bhattacharya, 2025; Mellado-Silva et al., 2020). The AI feedback discussed here is operationalized on three levels that are progressively personalized: simple feedback offers immediate truth and brief explanations without adjusting to individual performance (functionally similar to conventional automated feedback); adaptive feedback uses ML algorithms to adjust the type, depth, and format of feedback in real time by detecting repeated error patterns, modulating the complexity of explanations, and targeting identified weaknesses; and predictive feedback is built on an adaptive system with ML-based predictive models that anticipate future learning difficulties and implement proactive practices before they fail. The crucial difference is that temporal adaptive feedback is retrospective while predictive feedback is prospective. Previous research has shown ML-based feedback can improve conceptual understanding, motivation, and academic achievement by making teaching more responsive and individualized, which is especially valuable in accounting education where gradual problem-solving and error-prone calculations risk reinforcing misunderstandings (Adhikari & Shrestha, 2022; Gunarathne et al., 2020; Holtzblatt & Tschakert, 2011). Cognitive Load Theory provides theoretical reasoning: adaptive and predictive feedback reduces excessive cognitive load by facilitating problem-solving and freeing up working memory for schema construction and (Renkl & Atkinson, 2003; Skulmowski & Xu, 2022; Sweller et al., 2011; Van Merrienboer et al., 2003; van Nooijen et al., 2024). Based on this, the following hypotheses are formulated.

H1 (Main effect of AI-driven feedback): AI-based automated feedback has a significant positive effect on accounting learning outcomes, with adaptive and predictive feedback yielding higher results than simple feedback.

Learning Styles and Their Influence on Learning Outcomes

Learning styles show a relatively stable preference for understanding, processing, and storing information. Referring to Kolb's experiential learning theory (Kolb & Kolb, 2008; Lange & Kerr, 2013; Rhodes, 2013; Stewart & Khan, 2021), the study highlights three styles relevant to accounting education: visual learners who prefer diagrams and graphical representations, kinesthetic learners who prefer hands-on problem-solving and simulation, and convergent learners who combine abstract conceptualization with active experimentation, aligned with accounting needs for conceptual precision and procedural execution (Manolis et al., 2013; Eloff, 2017; Stanciu et al., 2020).

Accounting learning outcomes in this study were measured through cognitive dimensions (conceptual understanding and procedural problem-solving), practical (application of accounting concepts in realistic tasks), and affective (motivation, engagement, and attitude). Affective factors, which were poorly examined in previous studies, mediated the relationship between instruction and cognitive engagement because positive affectations increase sustained effort, intrinsic motivation, and the enhancement of long-lasting learning (Venter et al., 2025). In accounting education, where perceived difficulties can inhibit engagement, students' affective responses to feedback are critical; AI-based feedback has been shown to affect not only cognition but also motivation and engagement. Meta-analyses reported moderate effects of personalized AI feedback on learning outcomes and stronger effects on motivation, suggesting affective responses mediated cognitive achievement (Chang et al., 2023; Chen et al., 2021; X. Wang et al., 2022). Studies of empathic chatbots and explainable AI dashboards also show that affective and metacognitive feedback support self-directed performance and learning (Ortega-Ochoa et al., 2024; Skulmowski & Xu, 2022). Since affective responses and motivations vary according to individual characteristics and learning preferences, we hypothesize that learning styles will be associated with differences in overall accounting learning outcomes.

H2 (Main effect of learning styles): Learning styles have a significant influence on accounting learning outcomes, so visual, kinesthetic, and convergent students show different levels of overall performance.

The Interaction Between AI-Driven Feedback and Learning Styles

Although previous studies have independently documented the benefits of AI-based technologies and the relevance of learning styles, their co-effects in accounting education have rarely been examined. From a theoretical standpoint, the effectiveness of AI feedback likely depends on the alignment between the type of feedback and the processing mode students prefer. For visual learners, adaptive feedback that dynamically combines visual cues, step-by-step graphical explanations, and annotated solutions may be particularly effective. Kinesthetic learners may respond more positively to predictive feedback that leads them to interactive and simulation practice tasks that match their preference for hands-on iterative learning (Stewart & Khan, 2021). Convergent learners, who value the application of abstract concepts to concrete problems, may be expected to benefit from adaptive and predictive feedback, which provide directional cues, scenario-based exercises, and opportunities to test and refine understanding in applied contexts, as aptitude-treatment interactions, the direction and strength of such alignment are empirical questions.

Thus, AI-based feedback and learning styles are likely to interact in shaping accounting learning outcomes, so that the combination of certain levels of feedback and learning styles results in more favorable outcomes than others. Factorial experimental design is well-suited to capture the effects of such interactions by simultaneously changing the type of feedback and examining its effects across different categories of learning styles. Based on this, the following hypotheses are proposed:

H3 (Interaction effect): The effect of AI-based feedback on accounting learning outcomes varies across learning styles; There is a significant interaction between feedback levels and learning styles in determining overall learning outcomes.

METHOD

This study used a 3×3 factorial laboratory experiment between subjects consistent with an aptitude-treatment (ATI) interaction design (Cronbach & Snow, 1977). 1) AI-based feedback, with three levels (direct rule-based, personalized retrospective ML, and ML-predictive prospective), was experimentally manipulated, 2) learning styles, with three categories (visual, kinesthetic, and convergent), are measurable learner characteristics used to classify participants rather than manipulated variables. This design allows estimation of the main effects of feedback, the association of learning styles with outcomes, and their interactions. The nine cells generated are presented in Table 1.

Table 1. 3×3 Factorial Design: Experimental Conditions

Category		**AI-Driven Automated Feedback (U)**		
		Simple (S)	Adaptive (A)	Predictive (P)
Learning Style (G)	Visual (V)	Group 1 VS	Group 2 VA	Group 3 VP
	Kinesthetic (K)	Group 4 KS	Group 5 KA	Group 6 KP
	Converger (C)	Group 7 CS	Group 8 CA	Group 9 CP

Note. Each cell represents a unique experimental condition (n = 15 per cell). V = Visual; K = Kinesthetic; C = Converger; S = Simple; A = Adaptive; P = Predictive.

Participants are 145 undergraduate accounting students who are enrolled in the Intermediate Financial Accounting program at the University of Muhammadiyah Jember during the 2025/2026 academic year. This course was chosen because it requires conceptual reasoning (asset recognition and measurement) as well as procedural implementation (journal entry, depreciation calculation). All students provide informed written consent before participating and are informed that participation is voluntary, will not affect academic rankings, and involves an AI-assisted feedback system; Student data is anonymized prior to analysis. The study procedure is further described in the ethics statement.

Learning styles were established at Week 1 using the Kolb Learning Style Inventory 4.0 (KLSI 4.0), which resulted in four quadrant scores (divergent, assimilation, convergence, accommodation); each student's dominant quadrant determined the assigned category. Of the initial 145 participants, 10 students with a divergent dominant quadrant were excluded from the analysis, as the study theoretically focused on three profiles relevant to the accounting dual demands of conceptual representation and procedural execution; so the final sample was N = 135. The three preserved quadrants of KLSI were mapped to operational labels as follows: assimilation to visual, accommodation to kinesthetic, convergence to convergence. This renaming is based on functional alignment: assimilation learning involves abstract conceptualization elaborated through reflective observation, which operationally corresponds to visual preferences to representations and diagrams; accommodation learning involves active experimentation and hands-on experience, which corresponds to kinesthetic preferences to iteration and hands-on practice; while convergence retains its label because its integration between abstract conceptualization and active experimentation does not require remapping to sensory labels. Students whose quadrant scores are within a narrow range of each other are reviewed individually with secondary indicators before a final category is assigned. After the classification of learning styles, students are randomly assigned to one of three feedback conditions in their style blocks (stratified random assignments), resulting in a balanced allocation of 15 students per cell shown in Table 1.

The experiment lasted for three weeks in a university computer lab using AIFeed v2.1, a Moodle-based adaptive learning plugin with a large language model backend. In the immediate rule-based condition, the machine learning module is disabled and students receive truth feedback with brief explanations; in the ML-personalized retrospective condition, the ML module detects error patterns in each student's response history and generates feedback tailored to those patterns; in prospective ML-predictive conditions, the classifier also predicts possible areas of difficulty and implements proactive practices before anticipated errors (Figure 1).

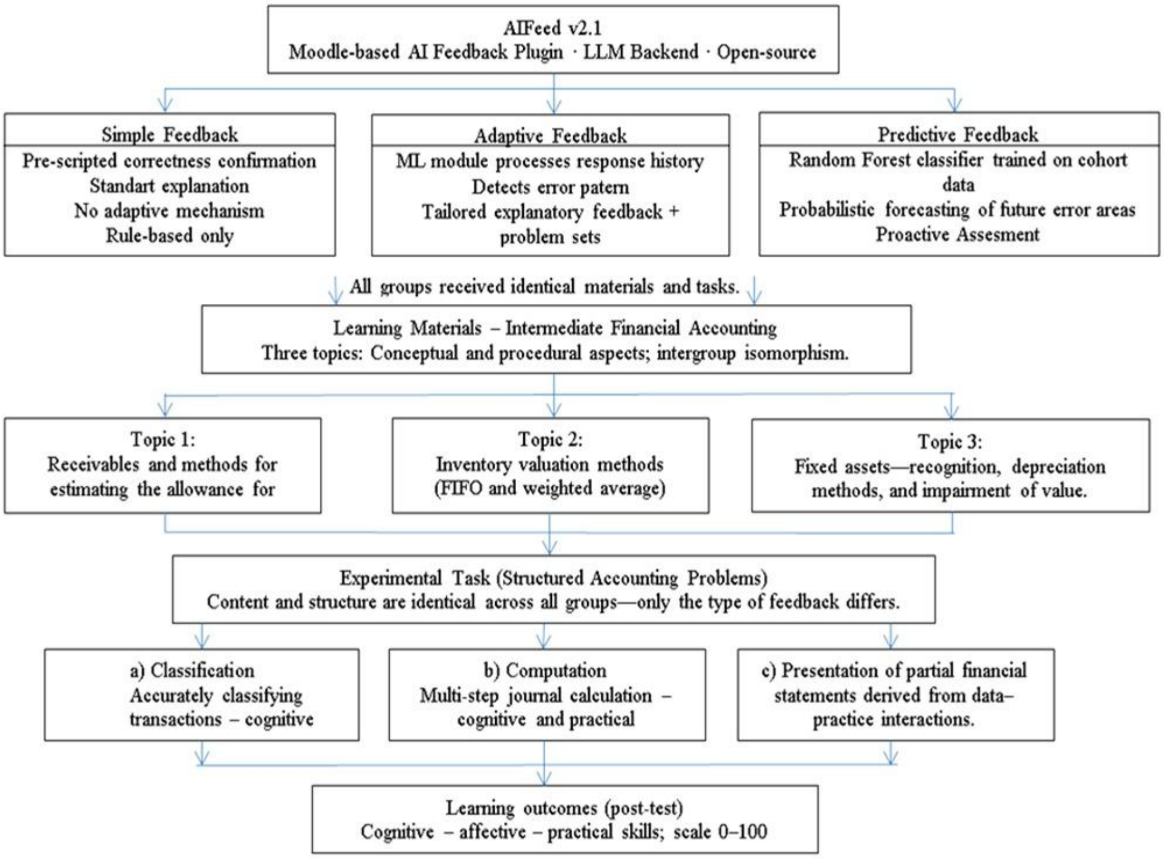

Figure 1. Treatment Structure of the AIFeed v2.1 Experimental Platform: Feedback Conditions, Learning Materials, and Task Types

Technically, AIFeed v2.1 uses a generative language model based on a large language model (LLM) as a feedback text generation engine; the ML retrospective component identifies recurring error patterns in the student's response history and uses it to adjust the type and depth of feedback delivered; while the predictive component adds a trained classifier to the student's initial response features (initial item error patterns and trial sequences) to anticipate areas of difficulty on subsequent assignments and implement proactive exercises. Treatment fidelity is verified through weekly system log audits that ensure each student only receives the type of feedback under the set conditions; no protocol violations were identified during the intervention period. Complete technical documentation regarding model versions and parameters, prompt design, and classifier validation procedures are managed separately by the development team as technical engineering reports.

The schedule of the experiment is as follows: Week 1 involves a pre-test (basic assessment) and the awarding of KLSI 4.0; The second week involves an instructional intervention, in which students work on an experimental task under established feedback conditions; Week 3 involves a post-exam. A willlike form is collected at the beginning of the experiment, as depicted in the procedure flowchart (Figure 2).

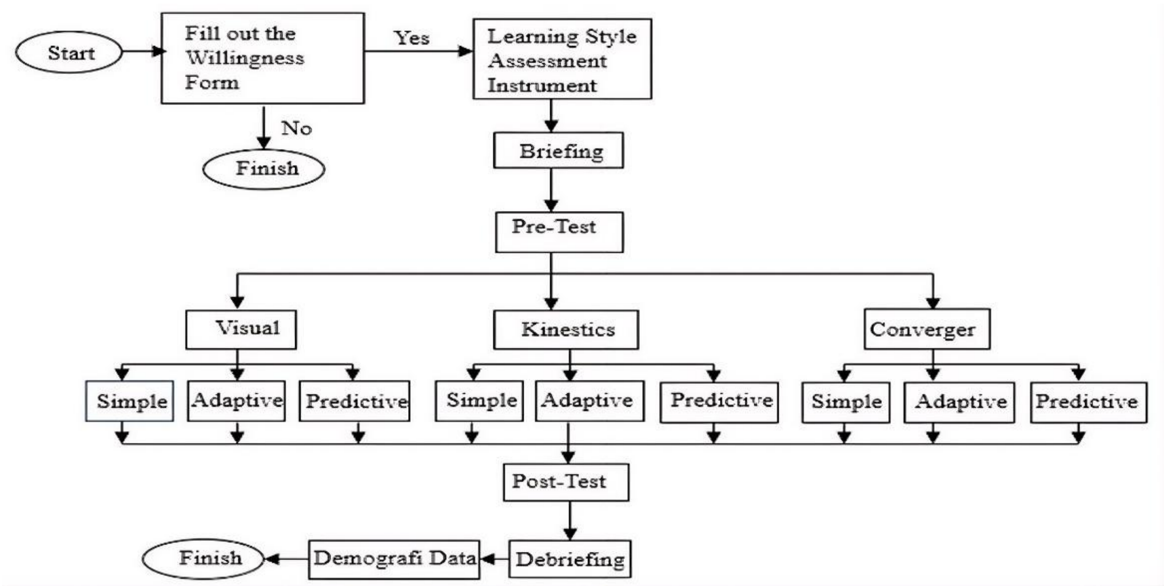

Figure 2. The full procedural flowchart.

The pre-test and post-test are constructed with a parallel and equivalent composite structure. The pre-test consists of three components that are structurally identical to the post-test: a cognitive sub-test (25 items of knowledge in an alternative-parallel form), an affective sub-scale (10 Likert items measure motivation and perceived values), and a rubric-based practical simulation, each graded on a scale of 0–100 and combined as an unweighted average. This structural equivalence ensures that pre-test and post-test scores can be compared directly, so that the gain score (post-test minus pre-test) meaningfully reflects changes in the construct of the same learning outcomes. Learning outcomes (post-test) were operationalized as composite scores that combined cognitive (25 post-test items, $\alpha = 0.83$), affective (10-item Likert scale measures motivation and perceived values, $\alpha = 0.79$), and practical (case simulation with rubric score, ICC = 0.87), each graded on a scale of 0–100 and aggregated as an unweighted average (Ebaid, 2022). The current analysis was performed on these combined results; Dimension-specific decomposition of cognitive, affective, and practical sub-scores was not available for the dataset underlying this analysis and was identified as a direction for future work in the conclusions.

The data were analyzed using two-way factorial ANOVA with AI-based automated feedback (Simple, adaptive, predictive) and learning styles (visual, kinesthetic, convergent) as fixed factors. The model can be expressed as:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (1)$$

where Yijk is the observed value for the combination of factors $_i$ and $_j$ and replication $_k$, μ is the overall average, αi is the effect of the AI-driven feedback factor at level $_i$, is the effect of the learning style factor at level $_j$, $(\alpha\beta)_{ij}$ is the effect of the interaction between the two factors, and εijk is a random error rate, which is assumed to be normally distributed with a mean of zero and constant variance.

Analytic Strategy

Data were analyzed by a predefined contingent analytical strategy. First, baseline equivalence among the nine cells was tested using bidirectional ANOVA on a pre-test composite score. Since the size of the primary outcome depends on whether baseline equivalence applies, the following rule is applied: if nine cells are equivalent at baseline, the post-test score will be the primary outcome; if not equivalent, the gain score (after the test is minus the pre-test) will be the primary outcome, with the analysis and analysis of the post-test covariance (ANCOVA, with the pre-test as the covariance) reported as an additional convergence check. For the primary model, residual normality is assessed by the Shapiro-Wilk test and the homogeneity of variance by the Levene test. The Tukey HSD post-hoc comparison is only calculated for effects for which the omnibus test is statistically significant, to avoid exploiting the coincidence of the unfiltered paired test set. Finally, the statistical strength achieved and the minimum detectable effect size at 80% power were calculated post hoc for the primary model using cohen's f. All analyses used $\alpha = 0.05$ (two tails).

Ethics Statement

This study has been reviewed and approved by the Institutional Research Ethics Committee of the University of Muhammadiyah Jember during the data collection; 83/II.3.AU/LPPM/Research/2026. In accordance with the institutional data governance protocol, students' interactions with AIFeed are recorded with an anonymous identifier, not a direct identifier, and students are informed during the approval process that feedback in two of the three conditions is generated by the AI-assisted system. Participation is voluntary, does not affect the trajectory position, and can be revoked at any time without penalty.

RESULTS AND DISCUSSION

Baseline Equivalence

Before evaluating the effects of treatment, the nine cells were compared based on a pre-test composite score (Table 2 reports a descriptive cell-level pre-test; Table 3 reports the corresponding ANOVA). Contrary to predictions of successful randomized tasks, pre-test scores differed significantly across different feedback conditions, with large effect sizes, and to a lesser extent across different learning style categories; Force-based feedback interactions on pre-test scores were not significant. The Levene test showed a heterogeneity of pre-test variance between cells.

Table 2. Descriptive Statistics of Pre-Test, Post-Test, and Gain Scores by Treatment Combination

Learning Style	**AI-Driven Automated Feedback**	**N**	**Pre-Test M (SD)**	**Post-Test M (SD)**	**Gain M (SD)**
Visual	Simple	15	63.66 (2.90)	74.65 (1.94)	10.99 (3.28)
Visual	Adaptive	15	67.05 (2.76)	79.13 (2.66)	12.08 (2.74)
Visual	Predictive	15	67.79 (1.49)	78.99 (3.12)	11.20 (3.26)
Kinesthetic	Simple	15	62.84 (3.73)	74.81 (2.28)	11.97 (4.29)
Kinesthetic	Adaptive	15	67.61 (2.99)	74.61 (1.68)	7.00 (3.36)
Kinesthetic	Predictive	15	68.56 (1.38)	79.49 (2.35)	10.93 (2.58)
Converger	Simple	15	64.01 (2.89)	76.10 (2.30)	12.09 (3.93)
Converger	Adaptive	15	68.86 (1.97)	77.67 (3.73)	8.82 (4.60)
Converger	Predictive	15	69.70 (2.44)	76.13 (3.96)	6.44 (4.93)

Note. M = mean; SD = standard deviation. Gain = post-test minus pre-test. Overall (N = 135): pre-test M = 66.68 (SD = 3.47); post-test M = 76.84 (SD = 3.29); gain M = 10.17 (SD = 4.18).

Table 3. Two-Way ANOVA of Pre-Test (Baseline) Composite Scores

Source of Variation	**SS (Sum of Squares)**	**df (degrees of freedom)**	**MS (Mean Square)**	**F Statistic**	**p (Sig.)**	**ηp^2**
Feedback	694.72	2	347.36	51.03	<0.001	0.448

Condition						
Learning Style	48.91	2	24.46	3.59	0.030	0.054
Feedback x Learning Style	15.05	4	3.76	0.55	0.697	0.017
Error	857.64	126	6.81	-	-	-
Total	1616.32	134	-	-	-	-

Note. ηp^2 = partial eta-squared. Levene's test on pre-test scores: F(8, 126) = 2.50, p = 0.015.

Because baseline equivalence does not apply, predefined analytical strategies establish learning improvement rather than post-test scores as the primary outcome, so that each student effectively serves as their own baseline control. Post-test analysis and adjusted covariates are reported afterwards as additional convergence checks, and the baseline imbalances themselves are examined further in the Discussion below as a limitation of causal interpretation.

Primary Analysis: Two-Way ANOVA on Learning Gain

Residual diagnostics support the use of ANOVA on gain scores: the Shapiro-Wilk test showed no deviation from normality (W = 0.981, p = 0.056), and the Levene test showed no heterogeneity of variance between cells (F(8, 126) = 1.43, p = 0.190), both of which were more favorable than diagnostics that were appropriate for both pre-test and post-test models. As shown in Table 4, the main effects of feedback conditions, the main effects of learning styles, and their interactions were all statistically significant.

Table 4. Two-Way ANOVA of Learning Gain (Primary Analysis)

Source of Variation	**SS (Sum of Squares)**	**df (degrees of freedom)**	**MS (Mean Square)**	**F Statistic**	**p (Sig.)**	**ηp^2**
Feedback Condition	155.85	2	77.93	5.56	0.005	0.081
Learning Style	122.51	2	61.26	4.37	0.015	0.065
Feedback x Learning Style	301.61	4	75.40	5.38	<0.001	0.146
Error	1,766.47	126	14.02	-	-	-
Total	2,346.44	134	-	-	-	-

Note. ηp^2 = partial eta-squared. Benchmarks: ≥ .01 small; ≥ .06 medium; ≥ .14 large (Cohen, 1988). Error df = ab(n − 1) = 9 × 14 = 126; Total df = N − 1 = 134.

The main effect of feedback conditions on gain was statistically significant (H1 partially supported), but not in the direction hypothetically anticipated: as Table 2 shows, the mean gain was numerically highest in simple (marginal M = 11.68) and lower in adaptive (M = 9.30) and predictive (M = 9.52) feedback. The main effect of learning style was also significant (H2 supported), with visual learners showing the largest average improvement and the least convergent learners. A significant force-based feedback interaction with a large effect size (H3 supported), suggests that the effect of feedback conditions on learning improvement is not uniform among learning style groups, a pattern examined in detail in the post-hoc comparison below.

Supplementary Convergence Analyses

Table 5 reports non-baseline-adjusted ANOVA on post-test scores and ANCOVA with pre-tests included as covariates, both conducted as convergence checks against the primary advantage score model.

Table 5a. Two-Way ANOVA of Post-Test Scores (Baseline-Unadjusted)

Source of Variation	**SS (Sum of Squares)**	**df (degrees of freedom)**	**MS (Mean Square)**	**F Statistic**	**p (Sig.)**	**ηp^2**
Feedback	210.72	2	105.36	13.74	<0.001	0.179

Condition						
Learning Style	39.92	2	19.96	2.60	0.078	0.040
Feedback x Learning Style	236.83	4	59.21	7.72	<0.001	0.197
Error	966.42	126	7.67	-	-	-
Total	1,453.89	134	-	-	-	-

Table 5b. ANCOVA of Post-Test Scores with Pre-Test as Covariate

Source of Variation	**SS (Sum of Squares)**	**df (degrees of freedom)**	**MS (Mean Square)**	**F Statistic**	**p (Sig.)**	**ηp²**
Feedback Condition	119.23	2	59.62	7.72	<0.001	0.110
Learning Style	39.75	2	19.88	2.57	0.080	0.040
Feedback x Learning Style	237.78	4	59.44	7.70	<0.001	0.198
Pre-Test (covariate)	0.97	1	0.97	0.13	0.724	0.001
Error	965.45	125	7.72	-	-	-
Total	1,363.18	134	-	-	-	-

On the post-test scores, the feedback and interaction conditions remained significant, and the learning style approached but did not reach significance; the marginal order of post-test feedback conditions (predictive and adaptive higher than moderate) corresponded to the direction initially anticipated on H1. This is consistent with the post-test comparison that only carried the basic advantages that the adaptive and predictive groups already had before the test (Table 2, Table 3), rather than reflecting treatment effects equivalent to the measures indicated by the advantage score model. ANCOVA reinforces this reading: once the feedback conditions, learning styles, and their interactions are included in the model, the pre-test scores essentially provide no additional explanatory power ($F(1, 125) = 0.13$, $p = 0.724$), suggesting that the pre-test differences are almost entirely absorbed by group membership rather than operating as an ever-changing, independent covariance. Feedback conditions, learning styles, and interactions all maintained a statistically significant and similarly sized effect on the ANCOVA specification, and mastery scores, which supported the reliability of the interaction findings themselves although they made it difficult to read purely causally to the main effects of the feedback.

Post Hoc Comparison

Because all three omnibus effects in the main advantage score model are significant, the Tukey HSD comparison is calculated both in each learning style group (comparing feedback conditions) and in each feedback condition (comparing learning style groups); results are reported in Table 6.

Table 6. Tukey HSD Post Hoc Comparisons on Learning Gain

Panel / Group	**Comparison**	**Mean Diff.**	**p (adj.)**	**Sig.**
Visual	Adaptive vs Predictive	0.88	0.717	ns
Visual	Adaptive vs Simple	1.09	0.603	ns
Visual	Predictive vs Simple	0.21	0.982	ns
Kinesthetic	Adaptive vs Predictive	-3.93	0.010	**
Kinesthetic	Adaptive vs Simple	-4.97	0.001	**
Kinesthetic	Predictive vs Simple	-1.04	0.693	ns
Converger	Adaptive vs Predictive	2.38	0.327	ns
Converger	Adaptive vs Simple	-3.27	0.128	ns

Converger	Predictive vs Simple	-5.65	0.004	**
Simple (feedback)	Converger vs Kinesthetic	0.12	0.996	ns
Simple (feedback)	Converger vs Visual	1.10	0.716	ns
Simple (feedback)	Kinesthetic vs Visual	0.98	0.766	ns
Adaptive (feedback)	Converger vs Kinesthetic	1.82	0.370	ns
Adaptive (feedback)	Converger vs Visual	-3.26	0.048	*
Adaptive (feedback)	Kinesthetic vs Visual	-5.08	0.001	**
Predictive (feedback)	Converger vs Kinesthetic	-4.49	0.005	**
Predictive (feedback)	Converger vs Visual	-4.76	0.003	**
Predictive (feedback)	Kinesthetic vs Visual	-0.27	0.979	ns

Note. ** $p < .01$; * $p < .05$; ns = not significant. Family-wise error was controlled within each panel using Tukey's honestly significant difference procedure.

The patterns in Table 6 are not a uniform story of "the more sophisticated feedback the better" as H1 anticipated, nor a simple analysis that one type of feedback is optimal for one style. For visual learners, gains did not differ significantly between a pair of feedback conditions: these groups obtained relatively good either simple, adaptive, or predictive feedback. For kinesthetic learners, adaptive feedback resulted in significantly smaller gains than simple and predictive feedback, while simple and predictive feedback did not differ from each other; In other words, the error-pattern-driven form of retrospective personalization was specifically associated with smaller improvements for this group, while the more basic, forward-looking forms of feedback did not. For convergent learners, the sequence is monotonous and the opposite of what CLT-based scaffolding arguments predict: the highest gain under simple feedback, the lowest under predictive feedback, with adaptive in between and statistically indistinguishable from any of the endpoints; only simple versus predictive contrast achieves significance.

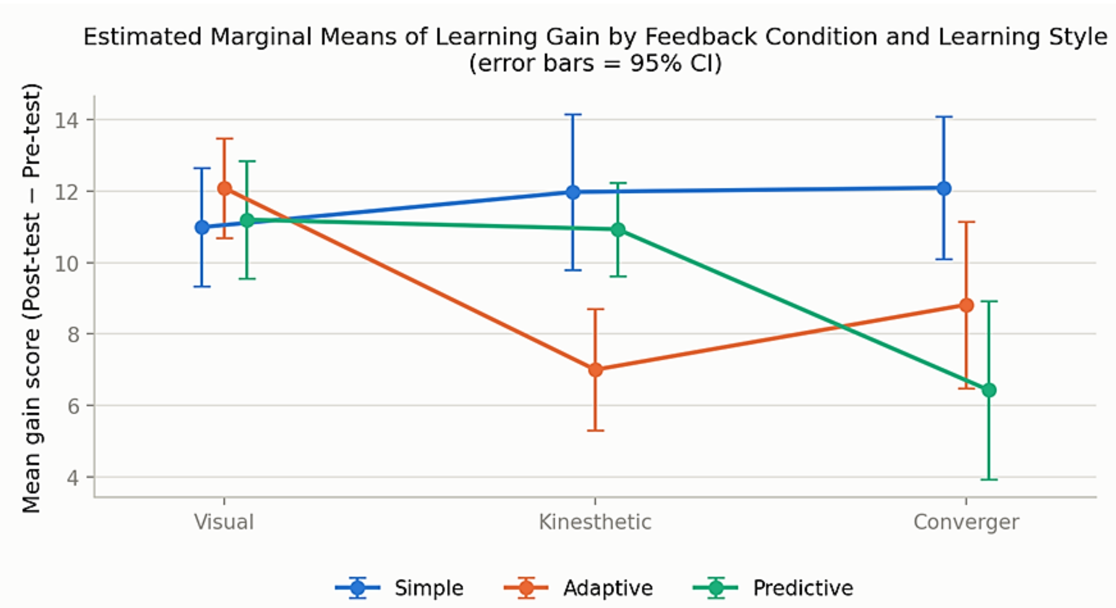

Figure 3. Estimated Marginal Means of Learning Gain by Feedback Condition and Learning Style (error bars = 95% CI)

When viewed based on feedback conditions, differences in learning styles in gain were not found in simple feedback (all paired comparisons were insignificant), meaning that all three groups of styles increased relatively when feedback was minimal. Under adaptive feedback, visual learners gained significantly more gains than kinesthetic learners and convergent learners. Under predictive feedback, visual and kinesthetic learners gained significantly more gains than convergent learners, respectively, while visual and kinesthetic learners did not differ from each other. Overall, interactions were concentrated on more sophisticated feedback conditions: simple feedback resulted in relatively flat and style-independent gain profiles, while adaptive and predictive feedback differed in effectiveness depending on learning style, and in two of the three groups of style divergence these were more favorable to less sophisticated conditions than more sophisticated ones.

Sensitivity and Statistical Power

The Table 7 report achieves the statistical strength for each effect in the primary model, calculated from the observed effect size.

Table 7. Sensitivity and Statistical Power Analysis (Primary Gain-Score Model)

Effect	ηp^2	Cohen's f	Achieved Power
Feedback (main effect)	0.081	0.297	0.856
Learning Style (main effect)	0.065	0.263	0.757
Feedback × Style (interaction)	0.146	0.413	0.975

Note. Power computed at $\alpha = .05$ using the noncentral F distribution given the observed effect size, error df = 126, and effect df = 2 (main effects) or 4 (interaction). At 80% power, the minimum detectable effect size would be $\eta p^2 \approx 0.071$ for a main effect and $\eta p^2 \approx 0.086$ for the interaction; all three observed effects exceed these thresholds.

With a balance of n = 15 per cell, this design is robust enough to detect the main effects and actually observed interactions, specifically the interactions (power ≈ 0.98); It will be much less capable of detecting smaller real effects than the intermediate range, a consideration relevant to the simple single-location sample discussed further below.

Integrating the Findings

The three observations, taken together, qualify how these results should be interpreted. First, these interactions are the most powerful and interpretively important findings here: they are significant and similar both estimated from the advantage score, post-test score, or ANCOVA, and describe a coherent, albeit unpredictable, pattern, visual learners are relatively less sensitive to the sophistication of feedback, kinesthetic learners are particularly disadvantaged by retrospective (adaptive) forms of personalization, and convergent learners get worse as the inputs become more sophisticated. Second, the main effect of feedback is more fragile: the marker depends on whether baseline differences are adjusted, which is also informative, suggesting that at least some of the adaptive and predictive feedback gains in posttest scores reflect pre-existing group differences, rather than feedback interventions alone. Third, and most importantly for how much causal weight the first two observations can bear, the pre-test groups are not equivalent at baseline despite the existence of a layered randomized assignment procedure intended to result in equivalence. This is a real threat to internal validity: it is impossible, with the available data, to fully separate the real effects of the feedback conditions from the pre-existing differences between students who end up in each condition. The interaction pattern is most likely not just a pure artifact of the baseline imbalance, as the force-based interaction on the pre-test score itself is insignificant (Table 3), the baseline imbalance primarily works on the primary effect of the feedback, not on the interaction, but this does not eliminate the need for caution, and future replication should audit and document the randomization procedure more rigorously than is possible here.

These findings indicate that: first, the sophistication of feedback helps some learners and not others, consistent with an aptitude-treatment interaction framework rather than a one-size-fits-all treatment (Cronbach & Snow, 1977). Second, for two of the three learning-style groups examined here, the more basic rule-based feedback conditions resulted in advantages that were statistically indistinguishable from, or numerically greater than, the more sophisticated ML-based conditions, an outcome that was not directly predicted by the CLT scaffolding argument, and which may reflect the additional processing demands imposed by more complex feedback (e.g., interpreting adaptive explanations or predictive practice recommendations) that were not equally accepted among the learning-style groups. This interpretation is given with caution, as the current data cannot judge between competing mechanisms; It is intended as a hypothesis for future work, not a fixed explanation.

The practical implications of configuring a single AI-based feedback that remains unlikely to serve all participants equally, which opposes a uniform implementation. At the same time, unresolved fundamental imbalances, single-institution and single-course samples, and reliance on composite outcome measures mean that a fixed automated feedback assignment policy based on learning style would be premature. A more sustainable interim measure would be for institutions to test AI-based feedback to maintain flexibility, allowing instructors to monitor how students respond and adjust feedback methods as needed, rather than committing to learning-based automated configurations prior to replication. Faculty development should also address how teachers can help students interact critically with AI-generated feedback rather than thinking of it as a substitute for independent reasoning, as highly targeted feedback systems can, if not monitored, serve as shortcuts to correct answers rather than real learning supporters.

CONCLUSION

AI-based feedback, learning styles, and interactions were associated with increased accounting learning on computationally intensive procedural tasks such as journal entries and depreciation schedules, but these patterns do not support the claim that more technologically advanced feedback is uniformly better. Visual learners gained comparable gains across all three feedback conditions; kinesthetic learners gained less retrospective (adaptive) feedback than simple or predictive feedback; and convergent learners gained less as feedback shifted from simple to adaptive to predictive. The most obvious and consistent contribution of this study was the interaction itself, which persisted across gain-score, post-test, and covariate-conforming specifications; The main effects of feedback conditions are more sensitive to how baseline differences are handled, and should be interpreted with appropriate caution.

Several limitations qualify the generalization and causal strength of these findings. Most importantly, pre-test scores differed substantially across different feedback conditions despite layered random assignment procedures intended to prevent this, which is a real threat to internal validity that cannot be fully resolved by the available data; gain-score and covariate-adjusted analyses were used to reduce, but not eliminate, these concerns. Samples ($N = 135$) were drawn from one institution and one course; a priori power calculation for this 3×3 design with a conventional medium effect (Cohen's $f = 0.25$) at $\alpha = 0.05$ and 80% power requires $N \approx$ 180–252 (depending on the assumed interaction value), so the sample size falls below such a priori requirement. However, as Table 7 shows, since the three observed effects were substantially large (Cohen's $f = 0.297$; 0.263; 0.413), the power achieved with $N = 135$ remained adequate for all three effects (achieved power = 0.856; 0.757; and 0.975). Thus, the power limitation is more relevant to the possibility of smaller real effects not being detected, rather than to the validity of the inference of the statistically significant effects reported here. The results were assessed at a single post-test time point, thus eliminating the conclusion about retention. The outcome measures used here are a combination of cognitive, affective, and practical indicators; Dimension-specific decomposition is not available for this analysis and may show a different pattern from the composites reported here. Individual differences such as previous academic achievement, digital literacy, and self-regulating learning capacity are uncontrollable. Finally, the tasks used in this study are procedural and computationally intensive; The interaction patterns identified here may not be generalized to other accounting subdomains, such as qualitative or ethical case analysis, that are not examined.

Future research should prioritize replication in which randomization and allocation procedures are documented and audited step-by-step, so that baseline imbalances can be diagnosed. In addition, replication with a larger multi-institutional sample, longitudinal follow-up to assess retention and transfer, and dimension-specific analysis of cognitive, affective, and practical sub-scores will each strengthen the trust or qualify for the interaction patterns identified here. The academic and metacognitive integrity dimensions of AI-assisted feedback in accounting education also deserve special attention in future experimental designs.

REFERENCES

Adhikari, D. R., & Shrestha, P. (2022). Knowledge management initiatives for achieving sustainable development goal 4.7: higher education institutions' stakeholder perspectives. *Journal of Knowledge Management*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/JKM-03-2022-0172

Andayani, E., Nurfaizana, D. R., Yudiono, U., Hariani, L. S., & Dameati, F. (2026). Virtual Economy Simulation Model (VESM): Interactive Learning Innovation in Introductory Economics Courses. *Journal of Economic Education and Entrepreneurship Studies*, *7*(1), 51–60. https://doi.org/10.62794/je3s.v7i1.21

Chaitali Bhattacharya. (2025). Enhancing Accounting and Business Management Education Through AI-Driven Personalized Learning. *Journal of Informatics Education and Research*, *5*(2), 1102–1111. https://doi.org/10.52783/jier.v5i2.2548

Chang, D., Lin, M., Hajian, S., & Wang, Q. (2023). Educational Design Principles of Using AI Chatbot That Supports Self-Regulated Learning in Education: Goal Setting, Feedback, and Personalization. *Sustainability*. https://doi.org/10.3390/su151712921

Chen, K., Huang, M., Zhu, X., & Wang, G. (2021). Learning disability early warning system based on classification algorithm. *2021 2nd International Conference on Information Science and Education (ICISE-IE)*, 597–600. https://doi.org/10.1109/icise-ie53922.2021.00141

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods: A handbook for research on interactions. Irvington. https://psycnet.apa.org/record/1978-11462-000

Ebaid, I. E. S. (2022). Sustainability and accounting education: perspectives of undergraduate accounting students in Saudi Arabia. *Journal of Applied Research in Higher Education*, *14*(4), 1371–1393. https://doi.org/10.1108/JARHE-05-2021-0183

Eloff, A.-M. (2017). The integration of information and information technology in accounting education: Effects on student performance. *Journal of Economic and Financial Sciences*, *9*(2), 409–425. https://doi.org/10.4102/jef.v9i2.49

Gunarathne, N., Senaratne, S., & Senanayake, S. (2020). Outcome-based education in accounting: The case of an accountancy degree program in Sri Lanka. *Journal of Economic and Administrative Sciences*, *36*(1), 16–37. https://doi.org/10.1108/JEAS-08-2018-0093

Haugland Sundkvist, C., & Kulset, E. M. (2024). Teaching accounting in the era of ChatGPT – The student perspective. *Journal of Accounting Education*, *69*(September 2023), 100932. https://doi.org/10.1016/j.jaccedu.2024.100932

Holtzblatt, M., & Tschakert, N. (2011). Expanding your accounting classroom with digital video technology. *Journal of Accounting Education*, *29*(2–3), 100–121. https://doi.org/10.1016/j.jaccedu.2011.10.003

Kolb, A. Y., & Kolb, D. (2008). Experiential Learning Theory: A Dynamic, Holistic Approach to Management Learning, Education and Development. 42-68. https://doi.org/10.4135/9780857021038.n3

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems. *Review of Educational Research*, *86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Lange, E. A., & Kerr, S. G. (2013). Accounting and incentives for sustainability in higher education: An interdisciplinary analysis of a needed revolution. *Social Responsibility Journal*, *9*(2), 210–219. https://doi.org/10.1108/SRJ-08-2011-0058

Mellado-Silva, R., Faundez-Ugalde, A., & Lobos, M. B. (2020). Learning tax regulations through rules-based chatbots using decision trees: A case study at the time of COVID-19. *Proceedings - International Conference of the Chilean Computer Science Society, SCCC*, *2020-Novem*. https://doi.org/10.1109/SCCC51225.2020.9281267

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology*, *114*(8), 1743–1772. https://doi.org/10.1037/edu0000764

Manolis, C., Burns, D. J., Assudani, R. H., & Chinta, R. (2013). Assessing Experiential Learning Styles: A Methodological Reconstruction and Validation of the Kolb Learning Style Inventory. Learning and Individual Differences, 23, 44-52. https://doi.org/10.1016/j.lindif.2012.10.009

Ortega-Ochoa, E., Quiroga Pérez, J., Arguedas, M., Daradoumis, T., & Marquès Puig, J. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies, motivation, self-regulation, and metacognitive reasoning in online higher education. *Internet of Things*, *25*, 101101. https://doi.org/10.1016/j.iot.2024.101101

Renkl, A., & Atkinson, R. K. (2003). Structuring the Transition From Example Study to Problem Solving in Cognitive Skill Acquisition: A Cognitive Load Perspective. *Educational Psychologist*, *38*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3

Rhodes, N. (2013). The Future of Accounting Education with the Integration of ICT. *University of Johannesburg (South Africa)*. https://hdl.handle.net/10210/12413

Skulmowski, A., & Xu, K. M. (2022). Understanding Cognitive Load in Digital and Online Learning: a New Perspective on Extraneous Cognitive Load. *Educational Psychology Review*, *34*(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7

Stanciu, V., Pugna, I. B., & Gheorghe, M. (2020). New coordinates of accounting academic education. A Romanian insight. *Journal of Accounting and Management Information Systems*, *19*(1), 158–178. https://doi.org/10.24818/jamis.2020.01007

Stewart, C., & Khan, A. A. (2021). A strategy for using digital mindsets and knowledge technologies to move past pandemic conditions. *Accounting Research Journal*, *34*(3), 345–356. https://doi.org/10.1108/ARJ-09-2020-0313

Suriyanti, Y., Sore, A. D., Relita, D. T., & Bobi, Y. (2026). Web-Based Learning Innovation in Economics Education : Leveraging Google Sites to Boost Interactivity and Conceptual Understanding. *Journal of Economic Education and Entrepreneurship Studies*, *7*(3), 33–46. https://doi.org/10.62794/je3s.v7i3.68

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer New York. https://doi.org/10.1007/978-1-4419-8126-4

Van Merrienboer, J., Kirschner, P., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist*, *38*, 5–13. https://doi.org/10.1207/s15326985ep3801_2

van Nooijen, C. C. A., de Koning, B. B., Bramer, W. M., Isahakyan, A., Asoodar, M., Kok, E., van Merrienboer, J. J. G., & Paas, F. (2024). A Cognitive Load Theory Approach to Understanding Expert Scaffolding of Visual Problem-Solving Tasks: A Scoping Review. *Educational Psychology Review*, *36*(1), 12. https://doi.org/10.1007/s10648-024-09848-3

Venter, J., Coetzee, S. A., & Schmulian, A. (2025). Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. *Assessment & Evaluation in Higher Education*, *50*(4), 516–536. https://doi.org/10.1080/02602938.2024.2415649

Wang, H., & Lehman, J. D. (2021). Using achievement goal-based personalized motivational feedback to enhance online learning. *Educational Technology Research and Development*, *69*(2), 553–581. https://doi.org/10.1007/s11423-021-09940-3

Wang, X., Zhang, L., & He, T. (2022). Learning Performance Prediction-Based Personalized Feedback in Online Learning via Machine Learning. *Sustainability*. https://doi.org/10.3390/su14137654

Wang, X., Huang, R., Sommer, M., Pei, B., Shidfar, P., Rehman, M., Ritzhaupt, A. D., & Martin, F. (2024). The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems From 2010 to 2022 on Learner Outcomes: A Meta-Analysis. Journal of Educational Computing Research, 62, 1348 - 1383. https://doi.org/10.1177/07356331241240459