International Journal of Contemporary Approaches in Education (IJCAE) 2025
Volume 4 (2) Page 81 - 100
https://doi.org/10.29329/ijcae.2025.1406.1

Human vs AI-Supported Feedback: Effects on Academic Achievement, Self-Regulation, and Feedback Literacy*

Dinçer Demir [1*], Sertel Altun [2], Ayfer Sayın [3]

Abstract

This study aimed to compare the effects of teacher feedback (TF) and AI-supported feedback (AIF) on academic achievement, perceived self-regulation, and feedback literacy among 42 sixth-grade students in a private school in Istanbul, Türkiye. Forty-two students were assigned to either a TF group (n=21), which received written feedback from the teacher, or an AIF group (n=21), which received AI-generated feedback through a Python-based natural language processing platform integrated with Cognitive Diagnostic Modelling. Both groups completed weekly quizzes over a four-week intervention period, aligned with English curriculum learning objectives. A 2 (time: pre-test vs. post-test) × 2 (group: TF vs. AIF) mixed-design multivariate analysis of variance (Mixed MANOVA) revealed significant improvements in all measured outcomes from pre-test to post-test ($p<.001$), with no significant differences between the TF and AIF groups or their interaction. These findings suggest that formative feedback enhances student outcomes regardless of delivery mode. The study underscores the potential of "AI + Teacher" collaborative models in middle school education, supporting essential skills development while addressing resource constraints for individualized feedback.

Keywords: Academic Achievement, AI-Supported Feedback, Feedback Literacy, Formative Feedback, Self-Regulation.

Received: 15.10.2025 – **Accepted:** 05.12.2025 – **Published:** 31.12.2025

* This study is derived from the doctoral dissertation prepared by the corresponding author in the Department of Curriculum and Instruction, Graduate School of Social Sciences, Yıldız Technical University.

[1] **Dinçer Demir,** Yıldız Technical University, Istanbul, Türkiye. ORCID: 0000-0003-4805-3593, dincherdemir@gmail.com
[2] **Sertel Altun,** Assoc. Prof. Dr., Yıldız Technical University, Istanbul, Türkiye. ORCID: 0000-0002-1951-5181, sertelaltun@gmail.com
[3] **Ayfer Sayın,** Prof. Dr., Gazi University, Ankara, Türkiye. ORCID: 0000-0003-1357-5674, ayfersayinsayin@gazi.edu.tr
[*] **Correspondence:** dincherdemir@gmail.com

INTRODUCTION

Feedback is a foundational element of the educational process, widely recognized by scholars as a powerful tool for enhancing learning outcomes and nurturing student development across various educational contexts (Bailey & Garner, 2010; Chickering & Gamson, 1987; Evans, 2013; Hattie & Gan, 2011; Hattie & Timperley, 2007; Wisniewski et al., 2020). It serves as a pivotal mechanism that directs students toward improved academic performance, fosters deeper comprehension of subject matter, and heightens engagement within classroom environments (Black & Wiliam, 1998). By delivering critical information about students' current progress and pinpointing specific areas requiring enhancement, feedback not only elevates academic achievement but also cultivates essential metacognitive skills, enabling learners to assume greater responsibility for their educational journeys (Nicol & Macfarlane-Dick, 2006).

Among the many forms of feedback, formative feedback stands out for its role in supporting ongoing learning. Defined as timely information that helps students adjust their learning strategies, it is highly effective due to its immediate and practical nature (Shute, 2008). Unlike summative feedback, which evaluates performance at the end of a unit, formative feedback enables students to adapt their approaches during the learning process (Sadler, 1989). Its effectiveness hinges on key attributes such as specificity in identifying strengths and weaknesses, goal-orientation aligned with learning objectives, and timeliness for prompt application (Brookhart, 2017; Hattie & Timperley, 2007). Research consistently demonstrates that formative feedback significantly enhances student achievement by promoting self-reflection and adaptive learning behaviors (Kluger & DeNisi, 1996; Morris et al., 2021; Wiliam, 2011).

Despite these proven benefits, providing timely, high-quality formative feedback poses a substantial burden on educators, who often manage extensive teaching responsibilities and large student cohorts (Carless, 2013; Carless et al., 2011; Hyland, 2019; Sandal et al., 2022). This resource-intensive process can overwhelm teachers, limiting their ability to deliver consistent, personalized support. To address these challenges, artificial intelligence (AI) has emerged as a groundbreaking tool with the potential to revolutionize feedback practices (Wongvorachan & Bulut, 2022). Recent advancements in AI integrate key technologies such as natural language processing (NLP), educational data mining (EDM), and learning analytics (LA), enabling systems to analyze student responses, identify patterns, and generate personalized feedback at scale (Wongvorachan & Bulut, 2022). Adoption has grown rapidly, with approximately 58% of K-12 educators reporting AI use in classrooms by early 2024 (Common Sense Media, 2024). These technologies are especially valuable in multiple-choice assessment contexts, transforming binary responses into meaningful learning opportunities through explanatory feedback (Wongvorachan & Bulut, 2022).

AI-supported feedback systems can rapidly analyze submissions, provide immediate responses, and tailor insights to individual needs, thereby alleviating teachers' administrative load (Luckin et al., 2016; Seo et al., 2021). Rather than replacing educators, contemporary research advocates for an "AI + Teacher" collaborative model, where AI generates initial feedback that teachers review, adapt, and enhance (Han & Li, 2024). This approach leverages AI's strengths in consistency, scalability, and pattern recognition while preserving human elements like contextual understanding, empathy, and pedagogical judgment (Wongvorachan & Bulut, 2022). For middle school students, who require structured guidance alongside relational support, this balanced model may be particularly beneficial (Anderman & Midgley, 1997; Pozdniakov et al., 2025). The versatility of AI is evident across educational domains: it enhances writing quality through automated mechanisms (Stevenson & Phakiti, 2014; Zeevy-Solovey, 2024), supports project-based learning in technical fields (Kusam, 2024), and delivers adaptive feedback to strengthen self-regulated learning (Mejeh et al., 2024). Recent studies show that generative AI tools like ChatGPT can produce high-quality, personalized feedback on writing, resembling expert human input (Steiss et al., 2024), while fine-tuned AI assistants yield promising results in K-12 settings (Castro et al., 2024).

Feedback profoundly shapes key outcomes such as self-regulation which is students' ability to monitor, control, and adapt cognitive processes to achieve goals and feedback literacy, which is the capacity to comprehend, interpret, and utilize feedback for improvement (Carless & Boud, 2018; Sutton, 2012; Zimmerman, 2000). Formative feedback fosters self-regulation by offering actionable insights into strengths and growth areas (Butler & Winne, 1995; Pintrich, 2000). Nicol and Macfarlane-Dick's (2006) model positions students as active agents generating internal feedback, which interacts with external sources, like teachers or AI, to refine interpretations, especially in middle school where metacognitive skills are emerging (Anderman & Midgley, 1997; Cutumisu & Schwartz, 2016; Pozdniakov et al., 2025). An ecological approach further emphasizes the interplay between feedback, learner receptivity, and the environment, viewing it as part of a continuous learning process rather than isolated events (de Kleijn, 2021). de Kleijn's (2021) instructional model outlines four activities such as seeking, making sense of, using, and responding to feedback that position students as active participants, underscoring the need for structured support in middle school.

Although AI's educational role has received significant attention, most research focuses on higher education (Zhai et al., 2021), leaving primary and secondary levels, which are particularly middle school, underexplored despite students' unique developmental needs (Crompton & Burke, 2022). Middle school marks a transitional phase toward independence, requiring scaffolded feedback for metacognitive development (Anderman & Midgley, 1997). This gap is acute in multiple-choice assessments (MCQs), a prevalent format criticized for promoting rote memorization over deep understanding (Bennett, 2011; Gierl & Lai, 2018; Thomas et al., 2025;). Evidence-based strategies, such as explanatory and comparative feedback (Nicol & Macfarlane-Dick, 2006) or indirect corrective approaches (Han & Li, 2024), can enhance MCQs' formative value, fostering cognitive engagement. Recent explorations show limited feedback in online MCQs boosts self-regulation (Say et al., 2024), while frequent assessments with feedback improve outcomes (Morris et al., 2021). For middle schoolers, feedback choices correlate with performance (Cutumisu & Schwartz, 2016), and this period is critical for feedback literacy (Pozdniakov et al., 2025). AI-supported systems offer transformative potential here, providing consistent, timely feedback as a complement to teacher guidance (Gao et al., 2024; Pozdniakov et al., 2025).

Given feedback's critical role in student learning and the resource challenges in middle school settings (Carless et al., 2011), this study is essential to investigate whether AI-supported feedback (AIF) can effectively complement teacher feedback (TF) in enhancing academic achievement, self-regulation, and feedback literacy among sixth-grade students. Middle schoolers' familiarity with digital technologies makes AI intuitive and engaging (Crompton & Burke, 2022), while AI's rapid evolution demands purposeful integration to optimize practices (Wongvorachan & Bulut, 2022). Through a quasi-experimental design with 42 participants in a private school, this research addresses key gaps, offering empirical insights into scalable feedback strategies that support middle school learners in developing vital skills for long-term success. Regarding the gaps addressed in the research, the following research questions have been identified:

1. When pre-test perceived self-regulation scores are controlled, is there a significant difference in post-test perceived self-regulation scores between the TF and AIF groups?

2. When pre-test feedback literacy scores are controlled, is there a significant difference in post-test feedback literacy scores between the TF and AIF groups?

3. When pre-test academic achievement scores are controlled, is there a significant difference in post-test academic achievement scores between the TF and AIF groups?

METHODOLOGY

Research Design

This study employed a quasi-experimental design to compare the effects of teacher feedback (TF) and AI-supported feedback (AIF) on academic achievement, self-regulation skills, and feedback literacy among 6th-grade students. A quasi-experimental design was selected due to practical constraints preventing random assignment in a school setting, enabling the comparison of two intact classes as experimental groups (Campbell & Stanley, 1963). The study comprised two groups:

Experiment 1 (TF group) received formative feedback from the teacher, while Experiment 2 (AIF group) received formative feedback from a fine-tunned AI system. Pre- and post-test measurements were conducted to assess the dependent variables which are academic achievement, perceived self-regulation, and feedback literacy while controlling for initial differences through the use of pre-test scores as covariates.

Study Group

The participants were 42 sixth-grade students (22 female, 20 male) enrolled in a private middle school in Istanbul, Türkiye. Two intact classes, both taught by the same English teacher, were randomly assigned to the TF and AIF groups (n = 21 per group). All students were native Turkish speakers learning English as a foreign language, with comparable prior achievement levels and socioeconomic backgrounds.

Data Collection Tools

Three instruments were utilized to collect data, administered as pre- and post-tests to measure the dependent variables:

Academic Achievement Test: A 20-item multiple-choice test was developed specifically for this study to evaluate students' academic achievement in English, aligned with the 6th-grade English curriculum and the unit theme "The Value of History - Identity" covered during the intervention. Initially, a specification table was created based on 13 learning outcomes across seven subtopics within the unit. A pool of 27 four-option multiple-choice items was drafted by two English subject experts and one assessment specialist. After expert review, 26 items were included in a pilot test administered to 284 seventh-grade students. Post-pilot item statistics were analyzed (Mean Item Difficulty: 0.607), and 20 items with an item discrimination index of 0.30 or higher were selected to ensure high content validity. The final test's reliability was established using the Kuder-Richardson 20 (KR20), yielding a value of 0.827, indicating good internal consistency.

Perceived Self-Regulation Skills Scale: Developed by Arslan and Gelişli (2015), this scale measured students' self-regulation abilities with 16 items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Higher scores reflected greater perceived self-regulation skills. The scale was implemented to 332 6th and 7th grade students before it was applied in the research and the scale's reliability in this study was found as Cronbach's alpha of 0.790, demonstrating acceptable internal consistency.

Feedback Literacy Scale: Developed by Yıldız et al. (2022), this scale assessed students' ability to understand, interpret, and use feedback, comprising 24 items rated on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Higher scores indicated greater feedback literacy. The scale was implemented to 327 6th and 7th grade students before it was applied in the research and the scale's reliability in this study was found as Cronbach's alpha of 0.922, reflecting excellent internal consistency.

Procedure

The study spanned a four-week period during the first semester of the 2024-2025 academic year (September–December 2024) at a private school in Istanbul, Türkiye. Prior to the intervention, both the teacher and the AI system were prepared to deliver formative feedback aligned with similar criteria to ensure comparability. The procedure unfolded as follows and showed in Figure 1:

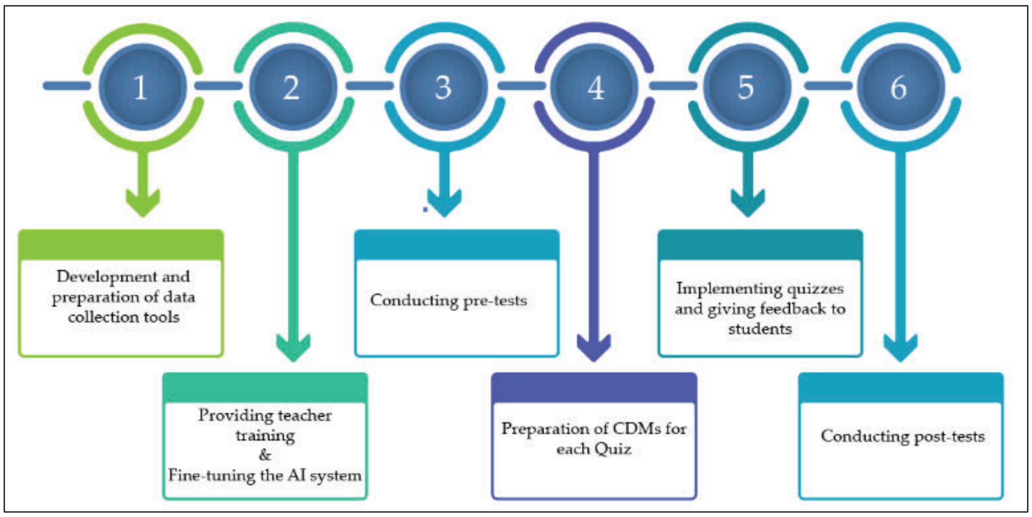

Figure 1. Procedure of The Intervention and Implementation

Preparation of Feedback Providers:

Teacher Training: The English teacher underwent a 2-hour online training session on formative feedback principles, focusing on its definition and effective delivery methods. The training, delivered by an educational specialist who is the researcher, emphasized providing feedback that was personalized (addressed to each student by name), timely (delivered within one to two days post-quiz), informative (detailing current performance), developmental (offering improvement strategies), and motivational (including encouraging statements). Prior to the study, the teacher conducted a pilot phase, delivering sample feedback on practice quizzes to five students of varying ability levels. These feedback samples were reviewed, and constructive feedback was provided to refine the teacher's approach, ensuring readiness for the intervention.

Design and Implementation of an AI-Supported Feedback System: The AI-supported feedback (AIF) system developed for this study was designed to deliver formative, individualized feedback to middle school students based on their performance in weekly multiple-choice quizzes. Rather than merely classifying responses as correct or incorrect, the system interprets student response patterns to diagnose underlying misconceptions and generates tailored feedback aligned with students' cognitive profiles. A cognitive model can be described as a structured representation of the knowledge, skills, and reasoning processes required to solve problems within a given domain. It provides a foundation for interpreting test performance by linking each item to the specific cognitive attributes that it is intended to measure (Gierl et al., 2021). Regarding the framework, the feedback system was composed of three integrated components that worked in sequence to identify misconceptions and guide improvement. The first component, the item-to-attribute mapping, associated each quiz item and its distractors with specific learning outcomes and cognitive features as shown in Figure 2. Distractor options were deliberately designed to reflect common misconceptions associated with learning outcomes. This mapping process was initially supported by rule-based tagging using pre-defined templates and then refined through expert validation to ensure diagnostic integrity. The overall item design and mapping strategy were informed by Sayın and Gierl's (2023) individualized feedback model, which integrates automated item generation with cognitive attribute tagging. GPT-4.0 was used to create the model. In the background, a test was entered for the completed model, and expert control was performed after the model was listed using Python. In the second stage, item-to-attribute mapping was matched with student answers. Students' strengths and characteristics to be developed in each subject are listed according to the map. Then, considering the feedback features outlined in the feedback training provided for the teacher, AI was reused to generate feedback for the students.

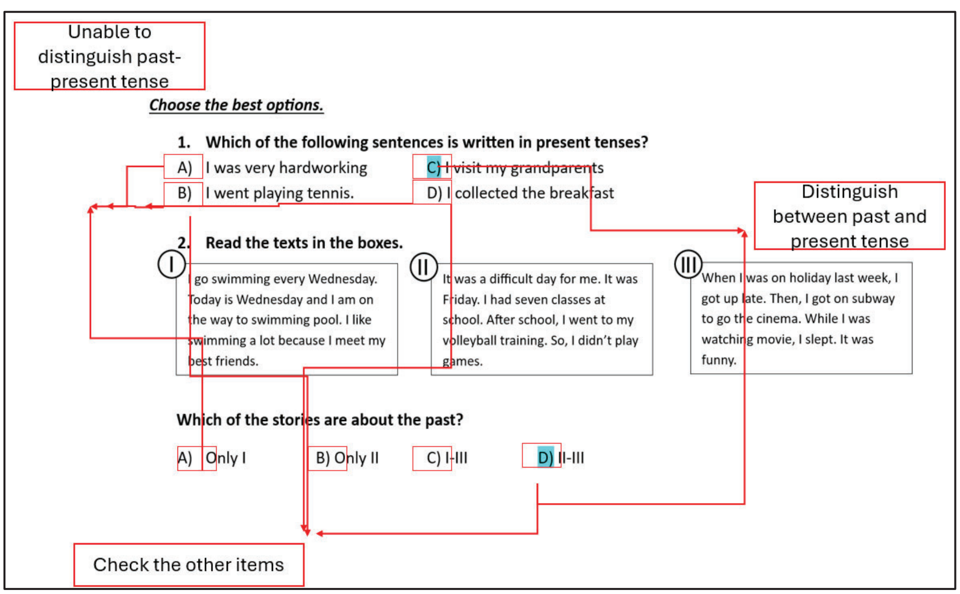

Figure 2. A Sample for the Item-to-Attribute Mapping

Pre-Test Phase: At the study's outset, both groups completed pre-tests—including the Academic Achievement Test, Perceived Self-Regulation Skills Scale, and Feedback Literacy Scale—in a single session during regular class hours to maintain consistency.

Intervention Phase: Four weekly quizzes, each consisting of 20 multiple-choice questions aligned with unit content and scored out of 20 points, were administered to provide opportunities for formative feedback. Over four weeks, the English teacher delivered identical unit content to both groups. Four weekly quizzes were administered, prepared by two English teachers to align with the unit's learning outcomes and the Academic Achievement Test. Each quiz targeted specific outcomes covered that week, with at least three questions per outcome to reduce chance factors. Post-quiz, the TF group received written feedback from the teacher on individual paper sheets (Figure 3), while the AIF group received AI-generated feedback on similar sheets (Figure 4). Feedback was printed by the teacher and distributed to students during class within one to two days of each quiz.

Sample 1:

Hi [Student's Name], in this quiz you got 10 correct and 10 incorrect out of 20 questions. I realize that you have improved compared to the previous quiz, and I congratulate you for your efforts. I found you a little too stressed in this quiz. As you continue your studies, your stress and the number of mistakes will decrease. I am sure of that. As for your mistakes, I would like you to remember that was/were is past tense. Remember that we use "While" when events continue for a while and "When" to describe specific moments, and I am sure that your success will increase again in the next quiz.

Sample 2:

Hi [Student's Name], 5 wrong answers in 20 questions are quite an achievement, congratulations! I am aware of your progress, and it makes me very happy. I can see the results that you are listening more attentively and actively to the lessons. Apart from that, I suggest you improve by asking me about our unit vocabulary and the words you don't know the meaning of in the activities during the lesson or by taking notes because 4 of your mistakes are due to this. For the other mistake, I suggest you read the text more carefully. I think your next quiz will be more successful and I expect you to continue like this.

Figure 3. Two Samples of the Feedback Given by The Teacher

Hello Dear Student [Student's Name],

Congratulations on completing the tracking test. I hope your hard work has paid off.

With a success rate of 80.00% you are doing great! I can see that you understand most of the topics very well. There are some small gaps, but with a little more work you can reach perfection. I will now provide you with some information that I think you might find useful.

Tense (Present/Past) Knowledge: I see that you know the Present Tense and how to distinguish tenses very well. You also know how to use the Past Tense, and you have made good conclusions about it by distinguishing tenses.

Question Patterns: I see that you understand the question patterns When, How Often, Who, Why exactly correctly.

Use of When While Conjunctions: You need to make a little more effort in using the past continuous tense with While. Although you generally understand the use of when and simple past tense, you can review this topic again.

Used To Structure: You know very well the structure of the phrase 'used to' and how to use it correctly. You know very well the structure of 'used to' for past habits.

Artificial Intelligence's Interpretation for You:

Ece, I am proud to see that you have an 80% success rate in the English tenses and question patterns test. I see that you have a wide knowledge of the subject and that you can use tenses correctly. Especially your successful performance in question patterns such as 'When', 'How often', 'Who', 'Why' and the 'Used to' pattern is remarkable. However, I see that you need to focus a little more on topics such as 'While + Past Continuous' and 'When + Simple Past'. I believe that you can take your success even higher by working extra on these. Maybe I won't be able to keep up with your speed either! Be confident, I am proud of you!

Until we meet again.

With love

Figure 3. A Sample of the Feedback Given by AI

Post-Test Phase: Following the intervention, identical post-tests were administered to both groups, enabling comparison of pre- and post-intervention scores across all dependent variables.

Data Analysis

Data Preparation

Prior to conducting the Mixed MANOVA, assumption checks were performed to ensure the validity of the analysis. Normality was assessed by examining skewness and kurtosis coefficients for each dependent variable in both groups, with all values falling within the acceptable range of ±2, indicating that the normality assumption was met (George & Mallery, 2010). Levene’s test confirmed homogeneity of variances across all dependent variables ($p > .05$). Outlier detection using z-standardized scores revealed no values exceeding ±3.29, indicating the absence of extreme outliers

(Tabachnick & Fidell, 2013). As the time factor consisted of only two levels (pre-test and post-test), the sphericity assumption was automatically satisfied (Field, 2013).

Analysis Procedure

To address the research questions, a 2 (time: pre-test vs. post-test) × 2 (group: TF vs. AIF) Mixed Multivariate Analysis of Variance (Mixed MANOVA) was employed to examine the effects of feedback type (Teacher Feedback [TF] vs. AI-Supported Feedback [AIF]) on academic achievement, perceived self-regulation, and feedback literacy, while accounting for changes over time. This approach was selected to assess both the main effects of time and group, as well as their interaction, across the three dependent variables simultaneously, considering their potential intercorrelations (Tabachnick & Fidell, 2013). The Mixed MANOVA was appropriate for the quasi-experimental design, allowing comparison of two intact groups while evaluating pre- to post-test changes, with pre-test scores serving as a baseline. The analysis was conducted using IBM SPSS Statistics (Version 26). The dependent and independent variables, their measurement methods, and the statistical analysis technique are summarized in Table 1.

Table 1. Summary of Variables and Analysis Procedure

Variable Type	Variable	Measurement Method	Statistical Analysis
Dependent	Academic Achievement	Academic Achievement Test (20-item multiple-choice test, scored out of 20 points)	Mixed MANOVA
Dependent	Perceived Self-Regulation	Perceived Self-Regulation Skills Scale (16 items, 5-point Likert scale)	Mixed MANOVA
Dependent	Feedback Literacy	Feedback Literacy Scale (24 items, 5-point Likert scale)	Mixed MANOVA
Independent	Time	Pre-test vs. Post-test (within-subjects factor)	Mixed MANOVA
Independent	Group	Teacher Feedback (TF) vs. AI-Supported Feedback (AIF) (between-subjects factor)	Mixed MANOVA

FINGINDS

To address the research questions, a 2 (time: pre-test vs. post-test) × 2 (group: Teacher Feedback [TF] vs. AI-Supported Feedback [AIF]) mixed-design multivariate analysis of variance (Mixed MANOVA) was conducted to examine the effects of feedback type on academic achievement, perceived self-regulation, and feedback literacy among 6th-grade students. The analysis assessed main effects of time and group, as well as their interaction, while controlling for pre-test scores as a baseline. Assumption checks confirmed that normality (skewness and kurtosis within ±2; George & Mallery, 2010), homogeneity of variances (Levene’s test, $p > .05$), and absence of outliers (z-scores within ±3.29; Tabachnick & Fidell, 2013) were met. Sphericity was automatically satisfied due to the two-level time factor (Field, 2013). Table 2 presents a summary of the 2 × 2 Mixed MANOVA results, including the main effects of time and group, their interaction, F-values, p-values, degrees of freedom, and partial eta-squared (η^2) effect sizes for all three dependent variables.

Table 2. Summary of Mixed MANOVA Results for Academic Achievement, Perceived Self-Regulation, and Feedback Literacy

Dependent Variable	*Effect*	*F (1, 40)*	*p*	*Partial η²*	*Interpretation*
Perceived Self Regulation	Time	34.72	<.001	.465	Significant main effect of time
	Group	0.23	.631	.006	No significant group effect
	Time × Group	0.26	.610	.007	No significant interaction
Feedback Literacy	Time	19.81	<.001	.331	Significant main effect of time
	Group	0.33	.568	.008	No significant group effect
	Time × Group	0.46	.503	.011	No significant interaction
Academic Achievement	Time	27.99	<.001	.412	Significant main effect of time
	Group	0.003	.960	.000	No significant group effect
	Time × Group	0.19	.662	.005	No significant interaction

Research Question 1: Differences in Self-Regulation Scores by Group

Table 3 displays the descriptive statistics for perceived self-regulation scores in both the Teacher Feedback (TF) and AI-Supported Feedback (AIF) groups at pre-test and post-test, including means, standard deviations, and normality indicators. The AIF group's post-test self-regulation mean was 57.81 (SD = 9.66), while the TF group's mean was 56.24 (SD = 8.47). Pre-test means were 53.86 (SD = 10.30) for AIF and 52.95 (SD = 8.62) for TF. Skewness and kurtosis values ranged between -1.22 and 0.14, confirming normality (George & Mallery, 2010).

Table 3. Descriptive Statistics for Self-Regulation Scores by Group

Measure	Group	N	Mean	Sd	Min	Max	Skewness	Kurtosis
Post-Test	AIF	21	57.81	9.662	43	73	-0.18	-1.22
	TF	21	56.24	8.467	38	67	-0.69	-0.58
Pre-Test	AIF	21	53.86	10.297	39	75	0.14	-0.90
	TF	21	52.95	8.623	36	65	-0.49	-0.68

When pre-test perceived self-regulation scores were controlled, the Mixed MANOVA revealed a significant main effect of time, $F(1, 40) = 34.72$, $p < .001$, partial $\eta^2 = .465$ (Table 2), indicating substantial improvement across both groups. However, no significant main effect of group was found, $F(1, 40) = 0.23$, $p = .631$, partial $\eta^2 = .006$, nor a significant time × group interaction, $F(1, 40) = 0.26$, $p = .610$, partial $\eta^2 = .007$ (Table 2). The AIF group's post-test mean was 57.81 (SD = 9.66), while the TF group's was 56.24 (SD = 8.47) (Table 3). These results suggest that both TF and AIF were equally effective in enhancing perceived self-regulation.

Research Question 2: Differences in Self-Regulation Scores by Group

Table 4 reports the descriptive statistics for feedback literacy scores across the TF and AIF groups at pre-test and post-test, with means, standard deviations, and skewness/kurtosis values to confirm normality. The AIF group's post-test feedback literacy mean was 89.38 (SD = 13.46), compared to 84.76 (SD = 13.96) for the TF group. Pre-test means were 84.90 (SD = 14.02) for AIF and 79.00 (SD = 18.74) for TF. Skewness and kurtosis values ranged between -1.16 and 1.28, indicating normality (George & Mallery, 2010).

Table 4. Descriptive Statistics for Feedback Literacy Scores by Group

Measure	Group	N	Mean	Sd	Min	Max	Skewness	Kurtosis
Post-Test	AIF	21	89.38	13.459	53	112	-0.86	1.28
Post-Test	TF	21	84.76	13.957	54	100	-1.07	0.22
Pre-Test	AIF	21	84.90	14.021	58	106	-0.46	0.50
Pre-Test	TF	21	79.00	18.740	40	107	-1.16	0.97

With pre-test feedback literacy scores controlled, a significant main effect of time was observed, $F(1, 40) = 19.81$, $p < .001$, partial $\eta^2 = .331$ (Table 2). Neither the main effect of group, $F(1, 40) = 0.33$, $p = .568$, partial $\eta^2 = .008$, nor the time × group interaction, $F(1, 40) = 0.46$, $p = .503$, partial $\eta^2 = .011$, was significant (Table 2). Post-test means were 89.38 (SD = 13.46) for AIF and 84.76 (SD = 13.96) for TF (Table 4), indicating comparable gains in feedback literacy across feedback types.

Research Question 3: Differences in Academic Achievement Scores by Group

Table 5 provides the descriptive statistics for academic achievement test scores in the TF and AIF groups at pre-test and post-test, including means, standard deviations, and normality diagnostics. The AIF group's post-test academic achievement mean was 14.43 (SD = 3.57), while the TF group's mean was 14.71 (SD = 2.85). Pre-test means were 11.95 (SD = 3.99) for AIF and 12.33 (SD = 3.01) for TF. Skewness and kurtosis values ranged between -1.08 and 0.37, confirming normality (George & Mallery, 2010).

Table 5. Descriptive Statistics for Academic Achievement Scores by Group

Measure	Group	N	Mean	SD	Min	Max	Skewness	Kurtosis
Post-Test	AIF	21	14.43	3.572	8	19	-0.40	-1.08
Post-Test	TF	21	14.71	2.849	9	20	-0.25	0.22
Pre-Test	AIF	21	11.95	3.993	4	18	-0.44	-0.63
Pre-Test	TF	21	12.33	3.006	8	18	0.37	-0.92

Controlling for pre-test academic achievement, a significant main effect of time emerged, $F(1, 40) = 27.99$, $p < .001$, partial $\eta^2 = .412$ (Table 2). No significant main effect of group, $F(1, 40) = 0.003$, $p = .960$, partial $\eta^2 = .000$, or time × group interaction, $F(1, 40) = 0.19$, $p = .662$, partial $\eta^2 = .005$, was found (Table 2). The AIF group achieved a post-test mean of 14.43 (SD = 3.57), compared to 14.71 (SD = 2.85) for TF (Table 5), confirming equivalent impact on academic achievement.

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

Discussion

This quasi-experimental study investigated the comparative effects of teacher feedback (TF) and AI-supported feedback (AIF) on academic achievement, perceived self-regulation, and feedback literacy among 42 sixth-grade students in a private middle school in Istanbul, Türkiye. The results, analyzed using a 2 (time: pre-test vs. post-test) × 2 (group: TF vs. AIF) Mixed Multivariate Analysis of Variance (Mixed MANOVA), revealed significant improvements in all three dependent variables, which are academic achievement, perceived self-regulation, and feedback literacy, from pre-test to post-test across both groups ($p < .001$), with large effect sizes (partial η^2 ranging from .331 to .465). However, no significant differences were observed between the TF and AIF groups, nor were there significant time × group interactions for any of the dependent variables. These findings indicate that both feedback modalities were equally effective in enhancing student outcomes, suggesting that the quality and structure of formative feedback, rather than its delivery source, are critical drivers of learning in middle school contexts.

The comparable effectiveness of TF and AIF aligns with Vygotsky's (1978) sociocultural learning theory, which posits that learning is optimized through structured guidance within students' zones of proximal development. Both feedback types provided personalized, timely, informative, developmental, and motivational feedback, meeting the criteria outlined by Hattie and Timperley (2007) for effective formative feedback. This structured guidance likely facilitated students' ability to engage with feedback, fostering academic and metacognitive growth irrespective of whether it was delivered by a teacher or an AI system. Cognitive load theory (Sweller, 2011) further explains these results, as both TF and AIF reduced extraneous cognitive load by delivering clear, focused feedback, enabling students to concentrate on essential learning processes.

The significant main effect of time across all dependent variables highlights the potency of formative feedback in promoting student development, consistent with prior research (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006). The absence of significant group differences suggests that the AI-supported feedback system, leveraging Cognitive Diagnostic Modelling (CDM) and Natural Language Processing (NLP), was able to mirror the diagnostic and developmental qualities of teacher feedback. This finding supports the "AI + Teacher" collaborative model advocated by Han and Li (2024), where AI's analytical strengths, such as consistency and scalability, are complemented by teachers' contextual understanding and pedagogical judgment. For middle school students, who are transitioning toward greater independence (Anderman & Midgley, 1997), the structured nature of both feedback types likely supported key feedback literacy activities which are seeking, making sense of, using, and responding to feedback, as outlined by de Kleijn's (2021) instructional model.

The improvements in feedback literacy are particularly noteworthy, as they align with Carless and Boud's (2018) conceptualization of feedback literacy as the ability to understand, interpret, and act on feedback. Both TF and AIF provided actionable insights that enabled students to engage actively with feedback processes, fostering their capacity to navigate feedback within their learning ecology (de Kleijn, 2021). The AI system's ability to generate personalized feedback addressing specific misconceptions, as enabled by CDM, likely enhanced its formative value, particularly in the context of multiple-choice assessments, which have been criticized for encouraging rote memorization (Bennett, 2011). The integration of CDM with automated item generation principles (Gierl & Lai, 2018; Sayın & Gierl, 2023) allowed the AIF system to transform multiple-choice quizzes into diagnostic tools, offering explanatory feedback that promoted conceptual understanding and metacognitive development, as suggested by Nicol and Macfarlane-Dick (2006).

Despite the equivalent outcomes, the lack of significant differences between TF and AIF may reflect the short duration of the intervention (four weeks), which may not have been sufficient for the AI system's advantages—such as scalability and rapid processing—to manifest distinctly. Teachers' feedback, enriched by contextual understanding and relational cues, may have resonated more strongly with middle school students, who often value motivational and empathetic elements (Carless & Boud, 2018). This aligns with findings by Erisyerico and Fauzan (2024), who noted that students perceive both AI and human feedback as effective when feedback quality is high, but human feedback may carry additional relational weight in certain contexts. Conversely, the AIF system's consistency and diagnostic precision, as evidenced by its CDM-driven approach, likely compensated for any lack of relational nuance, resulting in comparable outcomes.

The study's findings also address the research gap regarding AI applications in middle school education, as highlighted by Crompton and Burke (2022). While much of the existing literature focuses on higher education (Zhai et al., 2021), this study demonstrates that AI-supported feedback can be effectively implemented in K-12 settings, particularly for middle school students who are developing metacognitive skills and feedback literacy (Pozdniakov et al., 2025). The cultural context of a private middle school in Türkiye further underscores the applicability of AI-supported feedback across diverse educational settings, provided it is carefully designed and fine-tuned to align with curriculum goals and student needs. However, cultural responsiveness remains a critical consideration, as educational technologies are often developed in Western contexts and may require adaptation to local educational practices and values.

The results also have implications for multiple-choice assessments, which remain prevalent in educational settings despite their limitations (Thomas et al., 2025). Thes AI system's ability to provide detailed, misconception-targeted feedback challenges assumptions about the formative limitations of multiple-choice formats (Gierl & Lai, 2018). By integrating CDM, the AIF system offered feedback that was comparable to teacher-delivered feedback, supporting deeper cognitive engagement and independent problem-solving, as noted by Han and Li (2024). This suggests that AI can enhance the formative potential of multiple-choice assessments when designed with diagnostic rigor.

However, the findings contrast with studies like Er et al. (2024), which found teacher feedback to be more effective than AI-generated feedback in a higher education programming course. This discrepancy may be attributed to differences in educational context (middle school vs. higher education) and subject matter (English language learning vs. programming), highlighting the importance of tailoring AI feedback systems to specific developmental and disciplinary needs. Middle school students, who are accustomed to digital technologies (Crompton & Burke, 2022), may have found the AIF intuitive and engaging, contributing to its effectiveness in this study.

Ethical considerations, as emphasized by Akgun and Greenhow (2022), are also relevant. The AIF system was designed with transparency and teacher oversight, mitigating concerns about data privacy and algorithmic bias. However, broader implementation of AI-supported feedback would require ongoing attention to these ethical dimensions to ensure equitable and responsible use in educational settings.

Conclusion and Recommendations

This study demonstrated that both TF and AIF significantly enhanced academic achievement, self-regulation, and feedback literacy among sixth-grade students, with no significant differences between the two modalities. These results suggest that AI-supported feedback can be as effective as teacher feedback in enhancing student outcomes in middle school settings.

The theoretical implications extend to our understanding of how different feedback delivery methods influence learning processes. The findings support the view that the quality and content of feedback may be more important than the source. Both sociocultural learning theory and cognitive load theory provide useful frameworks for understanding why both feedback types were effective when they both provided appropriate scaffolding and reduced extraneous cognitive load.

The practical implications are substantial for educational settings where teacher resources are limited or where immediate feedback is beneficial. AI-supported feedback systems can potentially alleviate some of the burden on teachers while maintaining the quality of feedback provided to students. The "AI + Teacher" collaborative model offers a promising approach for leveraging technology to enhance feedback practices without sacrificing the human element of education.

However, it is important to note that AI-supported feedback should be viewed as a complement to, rather than a replacement for, teacher involvement. The role of the teacher remains crucial in designing effective learning experiences, interpreting AI-generated feedback when necessary, and providing the human connection essential for student motivation and engagement.

Future studies should investigate hybrid feedback models over longer periods to assess sustained impacts on student outcomes. Research should also explore AI feedback effectiveness across diverse subjects, assessment types, and educational levels, with a focus on qualitative insights into student perceptions. Ethical considerations, including algorithmic transparency and cultural responsiveness, should be prioritized to ensure responsible AI integration. Finally, developing pedagogical strategies to enhance students' feedback literacy in the context of AI-supported systems could further maximize their educational benefits.

Limitations

The study's small sample size (N=42) and single-school setting limit generalizability to other educational contexts. The four-week intervention may not have captured long-term effects, and the focus on multiple-choice assessments in English language learning restricts insights into other formats or subjects. The absence of qualitative data limits understanding of students' subjective experiences with TF and AIF. Additionally, cultural factors specific to the Turkish private school context may have influenced outcomes, necessitating caution when applying findings elsewhere.

REFERENCES

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431-440. https://doi.org/10.1007/s43681-021-00096-7

Anderman, E. M., & Midgley, C. (1997). Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. *Contemporary educational psychology*, *22*(3), 269-298. https://doi.org/10.1006/ceps.1996.0926

Arslan, S., & Gelişli, Y. (2015). Algılanan öz-düzenleme ölçeği: Bir ölçek geliştirme çalışması. *Sakarya University Journal of Education*, *5*(3), 67-74. https://doi.org/10.19126/suje.07146

Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in higher education*, *15*(2), 187-198. https://doi.org/10.1080/13562511003620019

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Brookhart, S. M. (2017). *How to give effective feedback to your students*. Ascd.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245-281. https://doi.org/10.3102/00346543065003245

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Ravenio Books.

Carless, D. (2013). Sustainable feedback and the development of student self-evaluative capacities. In *Reconceptualising feedback in higher education* (pp. 113-122). Routledge.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in higher education*, *36*(4), 395-407. https://doi.org/10.1080/03075071003642449

Castro, G. P. B., Chiappe, A., Rodríguez, D. F. B., & Sepulveda, F. G. (2024). Harnessing AI for Education 4.0: Drivers of Personalized Learning. *Electronic Journal of e-Learning*, *22*(5), 1-14. https://doi.org/10.34190/ejel.22.5.3467

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE bulletin*, *3*, 7.

Common Sense Media. (2024). *Generative AI in K-12 education: A white paper on current applications, future possibilities, and key considerations.* Retrieved from https://www.commonsensemedia.org/sites/default/files/research/report/generative-ai-in-k-12-education-white-paper-updated-aug-2024-final-2.pdf

Crompton, H., & Burke, D. (2022). Artificial intelligence in K-12 education. *SN Soc Sci* **2**, 113. https://doi.org/10.1007/s43545-022-00425-5

Cutumisu, M., & Schwartz, D. L. (2016). The Impact of Middle-School Students' Feedback Choices and Performance on Their Feedback Memory. *International Association for Development of the Information Society*.

de Kleijn, R. A. M. (2021). Supporting student and teacher feedback literacy: an instructional model for student feedback processes. *Assessment & Evaluation in Higher Education*, *48*(2), 186-200. https://doi.org/10.1080/02602938.2021.1967283

Er, E., Akçapınar, G., Bayazıt, A., Noroozi, O., & Banihashem, S. K. (2024). Assessing student perceptions and use of instructor versus AI-generated feedback. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13558

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, *83*(1), 70-120. https://doi.org/10.3102/0034654312474350

Erisyerico, M. L., & Fauzan, A. (2024). Comparing Students' Perceptions of AI and Human Feedback in Improving Writing Skills. *Scientia*, *3*(2).

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Gao, L., López-Pérez, M. E., Melero-Polo, I., & Trifu, A. (2024). Ask ChatGPT first! Transforming learning experiences in the age of artificial intelligence. *Studies in Higher Education*, *49*(12), 2772-2796. https://doi.org/10.1080/03075079.2024.2323571

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference* (10th ed.). Pearson.

Gierl, M. J., & Lai, H. (2018). Using automatic item generation to create solutions and rationales for computerized formative testing. *Applied psychological measurement*, *42*(1), 42-57. https://doi.org/10.1177/0146621617726788

Gierl, M. J., Lai, H., & Tanygin, V. (2021). *Advanced methods in auto-matic item generation.* New York: Routledge.

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, *126*, 103502. https://doi.org/10.1016/j.system.2024.103502

Hattie, J., & Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction* (pp. 263-285). Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hyland, K. (2019). *Second language writing*. Cambridge University Press.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological bulletin, 119(2), 254. https://doi.org/10.1037/0033-2909.119.2.254

Kusam, V. A. (2024). *Generative-AI assisted feedback provisioning for project-based learning in CS education* (Unpublished master's thesis). University of Michigan–Dearborn, Dearborn, MI.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education.* Pearson Education.

Mejeh, M., Sarbach, L., & Hascher, T. (2024). Effects of adaptive feedback through a digital tool–a mixed-methods study on the course of self-regulated learning. *Education and Information Technologies*, *29*(14), 1-43. https://doi.org/10.1007/s10639-024-12510-8

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), e3292. https://doi.org/10.1002/rev3.3292

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451-502). Academic Press.

Pozdniakov, S., Brazil, J., Mohammadi, M., Dollinger, M., Sadiq, S., & Khosravi, H. (2025). AI-Assisted Co-Creation: Bridging Skill Gaps in Student-Generated Content. *Journal of Learning Analytics*, 1-23. https://doi.org/10.18608/jla.2025.8601

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, *18*(2), 119-144. https://doi.org/10.1007/BF00117714

Sandal, A. K., Helleve, I., Smith, K., & Gamlem, S. M. (2022). Feedback practice in lower secondary school: Exploring development of perceptions of feedback practice among students and teachers. *Cogent Education*, 9(1), 2101236. https://doi.org/10.1080/2331186X.2022.2101236

Say, R., Visentin, D., Saunders, A., Atherton, I., Carr, A., & King, C. (2024). Where less is more: Limited feedback in formative online multiple-choice tests improves student self-regulation. *Journal of Computer Assisted Learning*, *40*(1), 89-103. https://doi.org/10.1111/jcal.12868

Sayın, A., & Gierl, M. J. (2023). Automatic item generation for online measurement and evaluation: Turkish literature items. *International Journal of Assessment Tools in Education*, *10*(2), 218-231. https://doi.org/10.21449/ijate.1249297

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *International journal of educational technology in higher education*, *18*, 1-23. https://doi.org/10.1186/s41239-021-00292-9

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., & Olson, C. B. (2024). *Comparing the quality of human and ChatGPT feedback of students' writing. Learning and Instruction, 91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51-65. https://doi.org/10.1016/j.asw.2013.11.007

Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International,* 49(1), 31–40. https://doi.org/10.1080/14703297.2012.647781

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

Tabachnick, B.G.,& Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Thomas, D. R., Borchers, C., Kakarla, S., Lin, J., Bhushan, S., Guo, B., Gatz, E., & Koedinger, K. R. (2025). Does multiple choice have a future in the age of generative ai? a posttest-only rct. In *Proceedings of the 15th International Learning Analytics and Knowledge Conference* (pp. 494-504). https://doi.org/10.1145/3706468.3706530

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Harvard University Press.

Wiliam, D. (2011). *Embedded formative assessment.* Solution Tree Press.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in psychology*, *10*, 487662. https://doi.org/10.3389/fpsyg.2019.03087

Wongvorachan, D., & Bulut, O. (2022). Feedback generation through artificial intelligence. In *The Open/Technology in education, society, and scholarship association conference* (Vol. 2, No. 1, pp. 1-9). https://doi.org/10.18357/otessac.2022.2.1.125

Yıldız, H., Bozpolat, E., & Hazar, E. (2022). Feedback literacy scale: A study of validation and reliability. *International Journal of Eurasian Education and Culture*, *7*(19), 2214-2249. http://dx.doi.org/10.35826/ijoecc.624

Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, *6*(3), 1482-1482. https://doi.org/10.29140/tltl.v6n3.1482

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J., Yuan, J. & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, *2021*(1), 812542 https://doi.org/10.1155/2021/8812542

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39). Academic Press