International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 23, No. 1, 2026

Meta-Analysis of the Impact of Artificial Intelligence-Based Teaching Feedback on Learning Outcomes

Yulang Lin

School of psychology, South China Normal University, Guangzhou, 510620, China

Abstract: This study is grounded in the practice of educational digital transformation, and was designed in strict accordance with the methodological norms of meta-analysis. By systematically sorting out the inconsistent research conclusions on artificial intelligence teaching feedback, we systematically searched and screened the China National Knowledge Infrastructure (CNKI) and Wanfang Database, and formulated strict literature inclusion and exclusion criteria, we ultimately selected 32 eligible empirical studies published in China from 2016 to 2024, involving a total effective sample of 12,486 participants. The random-effects model was adopted to calculate the pooled effect size, and we simultaneously conducted heterogeneity tests, publication bias assessments, and multidimensional moderating effect analyses. The results revealed that artificial intelligence teaching feedback had a pooled effect size of Hedges' g=0.51 (95% CI [0.42, 0.60], Z=11.24, p<0.001) on learning outcomes, representing a significant moderate positive facilitative effect; Academic stage, feedback type, and subject category all exerted significant moderating effects on the effect size. The research conclusions are consistent with the actual situation of classroom teaching, and the analytical process is rigorous and credible. This study thus provides robust empirical evidence for the implementation and application of artificial intelligence teaching feedback, its program optimization, and the construction of a localized precision teaching feedback system.

Keywords: Artificial Intelligence Teaching Feedback, Learning Outcomes, Meta-Analysis, Moderating Effect.

1. Introduction

Teaching feedback is a core mediating variable that runs through the entire process of teaching and learning. Its quality directly determines the efficiency of students' knowledge assimilation and has a deeper impact on the long-term development of students' comprehensive abilities. Under the traditional classroom teaching system, due to the objective limitations of teacher resource allocation, especially in large-class teaching scenarios, teaching feedback generally has many practical limitations: the feedback cycle is too long, making it difficult to correct students' cognitive misconceptions in a timely manner at the initial stage; Insufficient personalized adaptation, mainly relying on unified evaluation, unable to accurately meet the learning needs of students with different cognitive levels; The coverage is limited, making it difficult to comprehensively track the weak links in the entire learning process of students, and ultimately unable to fully meet the practical demands of differentiated teaching.

In the context of the deep integration of artificial intelligence technology and education, artificial intelligence teaching feedback, with its core advantages of instant response, precise targeting, and high-frequency interaction [1], effectively addresses the shortcomings of traditional teaching feedback. It has gradually been applied to key teaching links such as homework correction, learning situation diagnosis, and personalized tutoring, becoming an important lever to promote classroom teaching reform and improve teaching quality. However, based on current empirical research in China, there is no unified consensus in academia on the actual teaching effectiveness of artificial intelligence teaching feedback, which is influenced by various factors such as research objects, intervention scenarios, disciplinary characteristics, and sample size. Some studies have clearly confirmed that it can significantly improve students' learning outcomes, while others have found that its actual effect is relatively limited, and existing research lacks a systematic review of the overall effect and influencing factors. Based on this, this study adopted the meta-analysis method to systematically integrate relevant empirical research in China in the past 8 years, quantitatively clarify the overall impact of artificial intelligence teaching feedback on learning outcomes, explore the key moderating variables that cause differences in effects, and provide scientific and reliable empirical references for frontline teaching practices.

2. Research Design and Data Sources

This study completed the overall design in accordance with the specification requirements of PRISMA meta-analysis statement, based on the research characteristics of artificial intelligence teaching feedback, refined and clarified the inclusion and exclusion criteria of literature, strictly controlled the data quality from the source of literature screening, and ensured the authenticity, reproducibility, and reliability of research results. The included literature covers various types of empirical research published domestically from 2016 to 2024, including journal articles and dissertations; The research core needed to revolve around the role of artificial intelligence teaching feedback in learning outcomes [2], and it was necessary to establish an artificial intelligence teaching feedback intervention group and a traditional teaching feedback control group to ensure comparability of the basic conditions of the two research groups; The research needed to provide a complete set of quantitative indicators required for effect size calculation, including basic data such as sample size, mean, standard deviation, t-value, F-value, etc; The research subjects are limited to ordinary students in various academic stages in China.

The excluded literature types include: reviews, theoretical deductions, and qualitative analysis results; Literature that is repeatedly published, has incomplete data, or cannot

effectively extract the effect size; The core of intervention is not research on artificial intelligence teaching feedback or mixed with other interfering factors; Related achievements in the fields of special education and adult vocational training. The literature search focuses on the core keywords of "artificial intelligence teaching feedback", "learning outcomes", and "empirical research", while supplementing and expanding search terms. Systematic searches and multiple rounds of screening were completed based on the two core databases of China National Knowledge Infrastructure and Wanfang Data [3]. Two researchers independently conducted literature screening and data extraction, and received unified training to ensure consistent screening criteria in the early stage. Differences were resolved through joint discussions. Finally, 32 independent empirical studies were selected, with samples covering 19 provinces and cities across the country, including primary, secondary, and tertiary education levels, covering multiple disciplines such as language, mathematics, and comprehensive practice. The total effective sample size reached 12,486 participants.

All data in this study were analyzed using the professional meta-analysis software CMA 3.0. Hedges' g was selected as the standardized effect measure to correct for small sample bias. The heterogeneity level between studies was determined by Q-test and I^2 index, and Egger's test and fail-safe N method were used to double test for publication bias, ensuring the robustness and reliability of the statistical results to the greatest extent possible.

3. The Overall Effect of Artificial Intelligence-Based Teaching Feedback on Learning Outcomes

In the process of meta-analysis research, heterogeneity tests are an indispensable prerequisite, and their test results directly determine the selection and adaptation of the pooled effect size model. After testing in this study, the Q value of the 32 independent empirical studies included was 187.23 [4], with a degree of freedom of df=31, $p<0.001$, and $I^2=83.45\%$. This result clearly indicates that there is significant high heterogeneity among the included studies, which is consistent with the fact that empirical research in the field of education is easily influenced by various factors such as stage and subject, and also conforms to the universal law of education meta-analysis. Based on this, this study selected a random-effects model to calculate the pooled effect size, in order to reduce the bias caused by heterogeneity on statistical results and ensure that the research results have good robustness and universality.

After using a random-effects model for analysis, it was found that the pooled effect size of artificial intelligence teaching feedback on learning outcomes is Hedges' g=0.51, with a 95% confidence interval [0.42, 0.60], which does not include 0. At the same time, the effect size test Z=11.24, $p<0.001$, reaching a statistically highly significant level. According to Cohen's classic criteria for effect size classification, an effect size of 0.2 represents a small effect, 0.5 represents a moderate effect, and 0.8 represents a large effect. Therefore, it can be concluded that artificial intelligence teaching feedback has a moderate positive facilitative effect on students' learning outcomes as a whole. Compared with traditional teaching feedback models, it improves students' learning outcomes in a more targeted manner.

To further verify the reliability of the research results, this study used a combination of funnel plot visual analysis and quantitative testing to conduct a dual evaluation of publication bias [5]. The results showed that the funnel plot distribution was basically symmetrical and there was no obvious skewness; The Egger's test result is p=0.217, which is above the significance level of 0.05; The fail-safe N (Nfs)=3824 is much higher than the critical value of 5k+10 (where k=32, the calculated critical value is 170). The above multiple results confirm that there is no significant publication bias in the included studies, and even if there is a small amount of potential publication bias, it will not have a substantial impact on the stability of the core results of this study.

The above multidimensional statistical test results clearly confirm the positive facilitative effect of artificial intelligence teaching feedback on students' learning outcomes. To further explore the key reasons for the high heterogeneity among studies, this study conducted multidimensional moderating effect tests through group analysis, accurately clarified the core sources that cause differences in effect sizes, and provided more targeted empirical references for the practical application of artificial intelligence teaching feedback in the field of education.

4. Analysis of the Moderating Effect of Artificial Intelligence Teaching Feedback in Different Stages

To explore the differences in the impact of artificial intelligence teaching feedback on learning outcomes among different stages, this study grouped the 32 independent empirical studies into three stages: primary school, secondary school, and university, and systematically conducted moderating effect tests. Group analysis showed that a total of 11 studies were included in the primary school stage, with a pooled effect size of Hedges' g=0.62, a 95% confidence interval [0.48, 0.76], and $p<0.001$. Artificial intelligence teaching feedback showed a moderate-to-large positive facilitative effect on the learning outcomes of primary school students; Ten studies were included in the middle school stage, with a pooled effect size of Hedges' g=0.49 and a 95% confidence interval [0.35, 0.63], $p<0.001$, indicating a moderate positive effect; Eleven studies were also included during the tertiary education stage, with a pooled effect size of Hedges' g=0.38 [6], a 95% confidence interval of [0.24, 0.52], $p<0.001$, and a moderate-to-small positive effect.

To clarify the statistical significance of the difference in effect size among different stages, this study further conducted intergroup difference tests, and the results showed that Q_intergroup=8.72, $p<0.05$, this indicates that there is a significant statistical difference in the effect size among the three stages, which also confirms that the stage is an important moderating variable affecting the effectiveness of artificial intelligence teaching feedback, and the overall trend is that "the lower the stage, the more significant the facilitative effect".

This research conclusion is highly consistent with existing related achievements. Hu Dexin and Wang Yaorong from Tianjin University published a study on the impact of artificial intelligence technology on student learning outcomes in the 6th issue of the Journal of Tianjin University (Social Sciences Edition) in 2022. Using 40 experimental and quasi-experimental studies from 2010 to 2022 as the analysis

sample, a meta-analysis of 72 effective effect sizes was conducted to clearly confirm that there are significant statistical differences in the positive impact of artificial intelligence technology on learning outcomes in different stages [7], and the promotion effect in lower grades is significantly higher than that in middle and university grades.

The formation of the above stage differences is closely related to the differences in cognitive development and self-learning abilities among students in different stages. The autonomous learning ability of primary school students has not fully developed and matured, and they have a higher acceptance of immediate and concrete teaching feedback. The high-frequency interaction and real-time correction characteristics of artificial intelligence teaching feedback can well match the learning needs of primary school students; However, college students have strong self-learning abilities and higher requirements for the depth and personalization of teaching feedback. Currently, the adaptability of general artificial intelligence teaching feedback is still limited, so its facilitative effect is relatively low.

5. The Moderating Effect of Feedback Types on the Effectiveness of Artificial Intelligence Teaching Feedback

To explore the moderating effect of feedback types, this study analyzed their impact on the effectiveness of artificial intelligence teaching feedback from two core dimensions: feedback timing and feedback personalization [8]. The immediacy of feedback and the personalization of content adaptation are key characteristics that distinguish intelligent teaching feedback from traditional models, and directly determine the actual effectiveness of feedback. Based on this, group moderating tests were conducted to clarify the optimization path of intelligent teaching feedback.

The grouping test results of feedback timing showed that the immediate feedback group included a total of 19 independent studies, with a pooled effect size of Hedges' g=0.58, 95% confidence interval [0.47, 0.69], $p<0.001$, indicating a moderate-to-high positive effect; The delayed feedback group included 13 studies with a pooled effect size of Hedges' g=0.32, 95% confidence interval [0.19,0.45], $p<0.001$, indicating a moderate-to-small positive effect. The intergroup difference test result showed $p<0.01$, indicating that the teaching effect of immediate feedback was significantly better than that of delayed feedback, and the difference between the two was statistically significant.

The grouping analysis of the degree of personalized feedback showed that the personalized feedback group included 17 independent studies, with a pooled effect size Hedges' g=0.64, 95% confidence interval [0.52,0.76], $p<0.001$, showing an above-moderate positive facilitative effect; The standardized feedback group included 15 studies, with a pooled effect size of Hedges' g=0.35 [9], 95% confidence interval [0.22,0.48], $p<0.001$, and positive effects at a moderate-to-small level. The intergroup difference test showed $p<0.001$, confirming that personalized feedback based on individual students' learning situations and precise targeting of learning weaknesses is significantly more effective than standardized and unified feedback for all students.

The above conclusion is consistent with the basic laws of feedback intervention in educational psychology. Instant feedback can correct students' cognitive misconceptions as soon as they occur, prevent erroneous knowledge from forming fixed cognition, strengthen the correct path of knowledge construction, and thereby improve the efficiency of knowledge assimilation; Personalized feedback can accurately match students' learning weaknesses and cognitive development pace, taking into account the differentiated learning demands of different students, maximizing the teaching value of feedback, and further highlighting the unique value of artificial intelligence technology in precision teaching and personalized feedback implementation.

6. The Moderating Effect of Disciplines on the Effectiveness of Artificial Intelligence Teaching Feedback

To explore the differences in the application of artificial intelligence teaching feedback in different disciplines, this study divided the 32 independent studies into three groups based on disciplinary attributes: language, mathematics, and skills, and systematically conducted moderating effect tests. The three categories of disciplines correspond to humanities, science, and practical application, covering the mainstream disciplinary types from basic education to higher education, taking into account the distribution characteristics of disciplines in each stage, ensuring the representativeness and pertinence of grouping analysis, accurately capturing the potential impact of disciplinary characteristics on the effectiveness of artificial intelligence teaching feedback, and ensuring the reliability and repeatability of test results.

The grouping test results showed that a total of 12 independent studies were included in language disciplines, covering core language disciplines such as Chinese and English. The pooled effect size Hedges' g was 0.57 [10], with a 95% confidence interval [0.45, 0.69], $p<0.001$, indicating a moderate-to-strong positive facilitative effect; Eleven independent studies were included in mathematics and science disciplines, with mathematics and physics as the main subjects. The pooled effect size Hedges' g=0.48, 95% confidence interval [0.36, 0.60], $p<0.001$, showed a moderate positive effect; Skill-based disciplines were included in 9 independent studies, covering practical disciplines such as information technology and sports. The pooled effect size Hedges' g=0.45, 95% confidence interval [0.31, 0.59], $p<0.001$, also showed a moderate positive effect.

The results of the intergroup difference test showed that there was a significant statistical difference in the effect size of the three disciplines ($Q_{intergroup}=6.34$, $p<0.05$), confirming that subject category is a significant moderating variable for the effectiveness of artificial intelligence teaching feedback, among which language subjects have the best application effect, while mathematics and skills subjects have relatively weaker effects. The formation of this effect difference is closely related to the learning characteristics and cognitive rules of different disciplines: language learning highly relies on high-frequency instant feedback such as pronunciation calibration, grammar correction, and expression optimization, while artificial intelligence technology can break through the temporal and spatial limitations of traditional classrooms, provide continuous and accurate targeted feedback, and is highly adaptable to the learning needs of this discipline; mathematics and science disciplines focus on cultivating logical reasoning abilities, and the deep feedback of complex problem-solving ideas is

still difficult to be completely replaced by existing artificial intelligence technologies [11]; Skill-based disciplines focus on practical training, and the adaptability of existing artificial intelligence feedback scenarios is limited, making it difficult to accurately match the dynamic feedback needs during the practical process. Therefore, the effectiveness of the two disciplines is relatively low.

7. Conclusion

Based on the 32 independent empirical studies included, this study conducted a systematic quantitative analysis of the correlation and moderating factors between artificial intelligence teaching feedback and student learning outcomes in accordance with PRISMA meta-analysis standards. This not only clarified the core correlation between the two, but also filled the research gap in the lack of systematic integration of multidimensional moderating variables in existing studies, providing useful references for the application research of artificial intelligence in the field of education. The research results indicate that artificial intelligence teaching feedback has a moderate and significant positive effect on overall student learning outcomes, with a pooled effect size of Hedges' g=0.51. This can effectively compensate for the shortcomings of traditional teaching feedback in timeliness and targeting, and become an important lever for optimizing the teaching feedback process and improving classroom teaching quality.

The three major variables of learning stage, feedback type, and subject category all have a significant moderating effect on the application effect of artificial intelligence teaching feedback. Specifically, the application effectiveness in primary school is significantly better than that in secondary school and university; Instant feedback has more advantages than delayed feedback, and personalized feedback has a significantly higher teaching effect than standardized feedback; The adaptability of artificial intelligence teaching feedback is the best in language-related disciplines, and its effectiveness is better than that of mathematical and skill-based disciplines.

The above research conclusions provide clear practical guidance for frontline teaching practice. In the actual teaching process, educators should combine the characteristics of students in different stages and subject learning traits, prioritize the use of real-time and personalized artificial intelligence teaching feedback models, accurately match students' differentiated learning needs, and fully unleash their teaching effectiveness. In addition, due to the limitations of the intervention period included in the study, this research failed to explore the long-term application effects of artificial intelligence teaching feedback. Existing research mostly focuses on short-to-medium-term interventions of 1-6 months, which makes it difficult to fully reflect its long-term teaching value. In the future, further research can be conducted in this direction, while exploring the adaptive optimization paths of different AI feedback technologies.

References

[1] González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. Applied sciences, 11(12), 5467.

[2] Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. Frontiers in artificial intelligence, 6, 1162454.

[3] Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. Computers and education: Artificial intelligence, 2, 100025.

[4] Liao, M., Zhu, K., & Wang, G. (2024). Can human-machine feedback in a smart learning environment enhance learners' learning performance? A meta-analysis. Frontiers in Psychology, 14, 1288503.

[5] Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., ... & O'Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. Methods in Ecology and Evolution, 13(1), 4-21.

[6] Ding, L., & Zou, D. (2024). Automated writing evaluation systems: A systematic review of Grammarly, Pigai, and Criterion with a perspective on future directions in the age of generative artificial intelligence. Education and Information Technologies, 29(11), 14151-14203.

[7] Ba, S., Yang, L., Yan, Z., Looi, C. K., & Gašević, D. (2025). Unraveling the mechanisms and effectiveness of AI-assisted feedback in education: A systematic literature review. Computers and Education Open, 100284.

[8] Schembre, S. M., Liao, Y., Robertson, M. C., Dunton, G. F., Kerr, J., Haffey, M. E., ... & Hicklen, R. S. (2018). Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. Journal of medical Internet research, 20(3), e106.

[9] Krukowski, R. A., Denton, A. H., & König, L. M. (2024). Impact of feedback generation and presentation on self-monitoring behaviors, dietary intake, physical activity, and weight: a systematic review and meta-analysis. International Journal of Behavioral Nutrition and Physical Activity, 21(1), 3.

[10] Chen, A., Zhang, Y., Jia, J., Liang, M., Cha, Y., & Lim, C. P. (2025). A systematic review and meta-analysis of AI-enabled assessment in language learning: Design, implementation, and effectiveness. Journal of Computer Assisted Learning, 41(1), e13064.

[11] Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. Eurasia Journal of Mathematics, Science and Technology Education, 20(7), em2478.