SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

Comparing the Effects of AI and Traditional Answer Keys on Metacognitive Accuracy

By Hyunsu Song

AUTHOR BIOGRAPHY

Hyunsu Song is a junior at Korea International School located in Pangyo, South Korea. Observing the unique implications AI has on students in everyday life, he focuses on the similarities and differences in the influence of AI and traditional answer provision in shaping metacognitive monitoring. He is especially interested in educational applications of AI that balance performance with metacognitive accuracy.

ABSTRACT

How learners access correct answers can influence their ability to accurately assess their own understanding. Although traditional and AI answer provision were studied separately, no research directly compared their metacognitive effects. Using peer reviewed sources on metacognition and learning, this comparative literature review aims to analyze whether the source of correct answers differently impacts the relationship between the performance and metacognition of learners. By comparing traditional answer keys and AI assistance, this paper asserts that AI amplifies illusions based on fluency, causing learners to be overconfident in their learning. Additionally, it looks at the ability of traditional answer keys to mitigate the negative effects of quick answer provision through reflection. By comparing past answer provision techniques and present usage of AI, the paper highlights the importance of evaluating educational tools on metacognitive outcomes, understanding that the source of answer provision is a non-neutral variable. These findings suggest that educators must evaluate tools based not only on performance improvement but also their effects on accuracy of the students' self assessment.

Keywords: *metacognition, metacognitive accuracy, illusions of competence, fluency heuristic, answer provision, traditional answer keys, large language models, AI assisted learning, educational technology*

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

INTRODUCTION

Students have always relied on external sources for correct answers, whether it was through answer keys, books, teachers, or even tutors to verify their answers and understand their current educational performance. Providing students with answers clearly improves their performance since access to answers can help students to identify their errors and understand how they could improve. However, the issue is that while obtaining correct answers improves performance, the ease of accessing answers could create metacognitive illusions where students misunderstand their fluency for real understanding of content. Metacognition, often described as "Thinking about thinking" (Flavell, 1979) deals with one's own cognitive processes and is important in learning. Metacognitive monitoring skills allow individuals to track their progress, assess their confidence, and evaluate their learning process. When students can get answers easily, their metacognitive monitoring skills become compromised, as they manage to perform well on tasks while becoming worse at assessing what they truly know and do not know. This gap between individual performance and metacognition has become a critical concern as education technology continues to improve and new sources of answers readily become available for students to use (Fernandes et al., 2024; Bjork & Bjork, 2011).

Researchers have extensively studied how traditional methods of providing answers affect the performance and metacognitive accuracy of students. The research has revealed that traditional methods caused fluency-based illusions (Carpenter et al., 2013; Wang & Xing, 2019), but when they were enhanced with prompts for reflection, they can support accurate self-assessment (Tomanek et al., 2017). Separately though, recent research has examined the implications of AI assistance, particularly Large Language Models, and how they affect learning outcomes and metacognitive monitoring. Studies show that while AI assistance improves performance, users often overestimate their true understanding of the subject, demonstrating lower metacognitive accuracy compared to students working without AI usage (Fernandes et al., 2024). Despite increased research in this field, no studies have directly compared how traditional answer keys and AI assistance differently affect the performance metacognition relationship. This comparison matters in educational institutions because they increasingly use AI tools alongside traditional answer keys without understanding the impacts they have on metacognition of students. This paper focuses on this gap through a comparative literature review that examines how different sources of correct answers ultimately affect the students' metacognitive accuracies.

This paper examines how traditional answer keys and AI assistance differently affect the relationship between the performance of students and their metacognitive accuracy. The analysis draws on established theories on metacognitive monitoring, including topics such as fluency-based illusions and the Dunning Kruger effect, to understand how different answer sources affect the students' self-assessment of their abilities. This review shows that the source of correct answers impacts the performance outcomes and metacognitive processes

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

of students in different ways. This paper is organized into six sections, beginning with the theoretical frameworks (Section 2), examining traditional answer provision techniques (Section 3), analyzing AI effect on metacognitive accuracy (Section 4), providing comparative analysis (Section 5), and concluding with implications and future directions (Section 6).

THEORETICAL FOUNDATIONS FOR METACOGNITION

Normally, when students learn something new, they should be able to judge how well they have actually learned it. This ability is known as metacognitive accuracy, and it is more complicated than it seems at first glance. The core problem is that learners cannot access their cognitive states (Koriat, 1997), meaning they cannot simply look inside their minds to see the quality of their stored information. According to Nelson and Narens' (1990) framework of metacognition, widely regarded as the most dominant conceptual framework for metacognition, metacognition could be understood as having two levels. Specifically, Object-level cognitions deal with simpler tasks including encoding and information retrieval, while metalevel cognitions manage and observe the object-level processes (Nelson & Narens, 1990). There is a bidirectional relationship between the object level and the meta level, where "the information flowing from the meta-level to the object-level either changes the state of the object-level process or changes the object-level process itself" (Nelson & Narens, 1990, p. 127). The separation in these two levels is like having a supervisor who watches your learning process that adjusts your study approach by deciding what works and what does not. Yet, because learners lack direct access to their memory states, they must rely on various cues they have gathered to understand how well they have learned something (Koriat, 1997), which are pieces of information available during learning that they can use to gauge the quality of how well they learned something. Cues can come from various sources, whether it is the characteristics of the material itself, pieces of the learning experience, or understandings of what helps memory. The point is that accuracy of their judgments rely on whether learners are using cues that actually predict future performance: If they rely on incorrect or misleading cues their self-assessments would be wrong regardless of their confidence level (Koriat, 1997).

One of the main issues with metacognitive accuracy is that people commonly mistake their performance with actual learning, and this is what leads to what Bjork and Bjork call "illusions of competence" (Bjork & Bjork, 2011). This confusion is mainly caused by a core distinction between storage strength, how well a memory representation is interlocked with knowledge and skills, and retrieval strength, how accessible that memory representation is. This distinction is essential for performance as it is entirely dependent on retrieval strength, while storage strength could make up for losses in memory and act in the relearning lost retrieval strength (Bjork & Bjork, 1992). However, according to Bjork and Bjork, "the gain and loss of retrieval strength are both negatively accelerated" (Bjork & Bjork, 1992, p. 44), meaning that if learners confuse retrieval strength as their current storage strength, they might end up preferring worse learning conditions compared to better

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

ones. When trying to achieve better retention, difficulties in learning become preferred, as is described as "desirable difficulties" (Bjork & Bjork, 2011): Desirable difficulties include different types of learning such as interleaving, spacing, and tests, which are different from methods that do not use desirable difficulties like grouping instructions on topic, massing study sessions and using presentations as study events. (Bjork & Bjork, 2011). The confusion between retrieval strength and storage strength has been well documented across many studies. In a study by Rohrer and Taylor (2007), when students were asked to learn math formulas through either interleaved or blocked practice, interleaved students performed much better on the delayed test, with 63% accuracy achieved through interleaving and 20% achieved through blocking. Similarly, when Kornell and Bjork had students learn to identify painting styles through blocked or interleaved practice, 78% thought blocking was better. Yet, despite the feeling that they had learned better with blocking, their performance showed the contrary as interleaving yielded greater results (Kornell & Bjork, 2008). This reveals that the learners misjudged their performance during practice as proof of good learning, when in reality it was temporary retrieval strength, not real storage strength.

But why do learners fall for these illusions? The answer lies in a mental shortcut the brain uses called the fluency heuristic. Koriat's foundational work helps to explain this. When people rely heavily on processing fluency as a cue for assessing their understanding, they operate under the assumption that easily understood items are more likely to be remembered compared to those that are harder to learn (Koriat, 2010). According to dual process theory, theory-based judgements rely on the usage of belief on what helps memory, while experience based judgments rely on experiential cues found in task performance. Experience based metacognitive judgements rely on cues that are obtained through completing tasks (Koriat, 1997), and they may cause metacognitive feelings based on heuristics that occur without the brain even realizing it. As a result, learners could be influenced by fluency without even realizing it because the heuristic operates automatically and is not always consistent with what the person believes about learning. The fact that the fluency heuristic is caused through experience is proven through Koriat and Ackerman's finding that people often applied the heuristic in monitoring their own learning, but naturally applied it when predicting someone else's learning unless they've just experienced self-paced learning themselves (Koriat & Ackerman, 2010). This suggests that the fluency heuristic is mainly driven by experience, and is automatic rather than conscious, which makes it difficult for learners to recognise and override it when it causes issues in their learning.

While fluency is a powerful cue, it doesn't act alone as beliefs also play a critical role in metacognitive judgements. Through Wang and Xing's work, we know that fluency and prior beliefs contribute independently to metacognitive illusions when learning (Wang & Xing, 2019). Theory based judgements draw on learners' previous beliefs on what helps their learning, such as the common belief that blocked practices are better than interleaved practice when learning since it seems more organized. Explaining this phenomenon, Wang's research found that when participants entered a learning environment with the preexisting belief that blocking was better

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

than interleaving, they continued this belief even after they experienced that the interleaving practice yielded better results (Wang & Xing, 2019). This illusion only disappeared when researchers provided a clear explanation for why interleaving was better than blocking (Wang & Xing, 2019), showing that beliefs could be very resistant to change even when contradictory data is present. When fluency and beliefs align, the metacognitive illusion could become especially strong, but when conflicted, learners could tend to rely more on their beliefs (Wang & Xing, 2019).

Yet, not all fluency is misleading and not all disfluency helps learning, meaning that understanding when fluency is a valid cue and when it creates illusions is important. According to Weissgerber and Reinhard's research on boundary conditions, the timing of the test matters greatly for whether disfluency benefits learning (Weissgerber & Reinhard, 2017). They found that disfluent conditions had enhanced long-term retention of information when tested after a delay, but they did not benefit retention when tested immediately, which fits the distinction between performance and learning (Weissgerber & Reinhard, 2017). Additionally, the type of disfluency also matters as for disfluency to be desired, learners need understanding of background information and skills to adequately respond to a challenge, as otherwise disfluency just becomes a barrier to learning rather than a productive difficulty (Bjork & Bjork 2011; Weissgerber & Reinhard, 2017). Nonetheless, fluency could still be a valid cue; for example, remembering "concrete words are considered easier to process ... than abstract words", and "coherent text is considered easier to process and ... than incoherent text," so fluency signals better learning in these cases (Carpenter et al., 2013). Fluency is a valid cue when it stems from better encoding, but it's an invalid cue when it stems from surface level processing ease that does not translate to real information retention (Bjork & Bjork, 2011; Carpenter et al., 2013).

TRADITIONAL ANSWER PROVISION EFFECTS ON METACOGNITION

Before examining AI's effects on metacognition, we must first look at how traditional forms of answer provision, like answer keys, affect students' metacognitive accuracy. Traditional answer provisions include answer keys, feedback mechanisms, and worked examples. Research established that these tools improve objective performance of students, but their effects on metacognition are much more complex. According to Tomanek, enhanced answer keys could help students to understand how to regulate their own learning and use self-generated feedback, (Tomanek et al., 2017), which could lead to greater success in school. However, while these traditional methods could improve performance, they can also create fluency-based illusions that impair students' ability to accurately access their metacognitive accuracy. Establishing these baseline effects can become the foundation for comparison to AI's effects on metacognition.

Enhanced answer keys, as researched by Tomanek et al., goes beyond simple correct and incorrect answers by providing explanations, rationales for grading, and prompts for reflection. The initial problems of

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

student behaviors that came with answer keys was that they used answer keys to look for incorrect answers, as they believed that a correct answer indicated they understood the subject (Tomanek et al., 2017, p.7). One student even said that "I feel like if I got it right I probably knew what was on the enhanced answer key and stuff so there's not really a reason to go over it" (Tomanek et al., 2017, p. 7). This behavior reveals a metacognitive problem, as students were using the correctness of their answers to gauge their understanding, missing crucial opportunities to deepen their learning even when they got their answers right. To solve these problems, researchers added reflection questions and instructions to prompt their students to think about what made their answers correct, and how the concepts connected to broader learning topics. After the intervention, most students reported using the enhanced answer keys and considered using them in the future (Tomanek et al., 2017). This proved that instruction is necessary when students are exposed to enhanced answer keys as there was a difference in the way interviewed and non-interviewed students utilized the enhanced answer keys (Tomanek et al., 2017), as some students failed to utilize the answer keys to their full metacognitive potential.

Fluency, how easy it feels for information to be processed, is an important cue that students could use to assess their learning and understanding, but it could also be misleading. According to Wang, "The fluency heuristic is one of the most commonly used in both remembering and nonremembering tasks" (Wang & Xing, 2019, p. 101). For example, in the Carpenter lecture fluency study, the same instructor provided content in an educational setting, once in a fluent manner and once in a disfluent manner. The researchers found that learners believed they learnt better from the more fluent instructor over the disfluent one (Carpenter et al., 2013). Nonetheless, lecture fluencies did not affect the amount of information learned for students. Taken together, these results suggest that presentation style, not mastery of content, drives students learning confidence.

Fluency-based misjudgments occur specifically when students study using answer keys, and are often characterized as part of a broader phenomenon called "illusions of competence." The core mechanism that causes students to have inflated fluency misjudgements "stems from a fundamental difference between the conditions of learning and the conditions of testing" (Koriat & Bjork, 2005, p.187). Specifically, when students have access to both answer keys and questions at the same time, one's assessment of performance in the future will occur with an answer, causing a perspective bias. These perspective biases could be mitigated by imagining oneself later without the answer present, mentally discounting what one currently knows (Koriat & Bjork, 2005, p.187). For example, Newton (1990, as cited in Koriat & Bjork, 2005, p. 187) found in his "tapper" study, 50% of tappers, people that tap out the rhythm, estimated that listeners, those who listen to the taps, believed their listeners would be able to identify the song they were tapping out. In reality, only about 3% of listeners managed to identify what song was being tapped (Newton, 1990, as cited in Koriat & Bjork, 2005). This suggests that when students check their work with answer keys, they experience the same curse of knowledge where the answer keys make understanding content much easier than normal.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

Research on desirable difficulties reveals a paradox, where desirable difficulties are detrimental to short-term cognitive performance but beneficial to long-term learning. According to Bjork, although difficult conditions for learning may not always enhance short term learning, they could benefit conditions that aid long term learning (Bjork, 1994). The issue with answer keys is that they provide instant feedback to questions, which reduces difficulty and allows fluency, minimizing deeper learning. This issue becomes significant when combined with the fact that if learners interpret their retrieval strength as storage strength, they "become susceptible to preferring poorer conditions of learning to better conditions of learning" (Bjork & Bjork, 2011). This leads to a performance-metacognition gap where students with answer keys perform better, but it comes at the cost of their ability to understand and assess the quality of what they truly learned. Understanding this gap provides the baseline for understanding if AI assistance creates similar effects on learners' metacognition.

THE DUAL EFFECTS OF AI ON METACOGNITION

While both AI assistance and traditional answer provision create metacognitive challenges, AI assistance has the potential to create more severe versions of the same metacognitive problems. Recent work by Fernandes et al. (2024) demonstrates a clear pattern where AI support improves students' performance in tasks while undermining learners' ability to assess their learning accurately. In a study designed to assess AI LLMs effects on overestimation, participants used ChatGPT-4o to complete logical reasoning problems from the Law School Admission Test. In their study, participants who used large language models generally performed better yet showed increased overconfidence compared to those who completed tasks without AI assistance. This once again suggests that AI is associated with a performance and metacognition gap where tools improve performance whilst reducing learners' awareness of their actual performance. Fully understanding this gap requires understanding the details of both the benefits of performance and metacognitive costs.

The Fernandes et al. study provided evidence for the gap between the improvement of actual performance of students and their perception of their individual improvements. The study showed that participants who used ChatGPT-4o scored higher on LSAT-style questions than in the group that solved problems without AI (M=12.98 vs. M=9.45), indicating a clear improvement in performance (Fernandes et al., 2024, p.7). Between the two groups of study, there was approximately a 3.5 point difference in score from M=12.98 to 9.45. This result, however, was accompanied by a decrease in the metacognitive accuracy of learners, meaning they overestimated their performance. On average, AI assisted participants estimated they had around 16.5 out of 20 questions correctly, when they actually overestimated their performance by around 4 points in reality (Fernandes et al., p. 7). This indicates that there is an issue; students overestimated their performance, guessing that their performance was greater than their actual improvement (4 points of overestimation vs 3.5 points). This creates a problem where the illusion of competence is greater than the reality of the competence gained through AI usage.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

The common thread between AI usage and traditional answer key usage is that they both operate through the same fluency-based mechanisms, where the ease of obtaining answers to problems leads to false senses of mastery. Yet, when students use AI, the interaction is much smoother and answers appear clearly and quickly, creating high processing fluency that leads students to misinterpret as their own performance/understanding. This is supported by the fact that students view the quality of their own learning depending on the fluency of the lecture provided by instructors, and not actual learning (Carpenter et al., 2013). This reveals that AI could actually amplify the effect of fluency compared to traditional answer key methods since AI provides not only answers but detailed explanations that feel comprehensive, causing students to have a greater illusion of understanding. This point is paralleled by what Fernandes describes as a shift from reflective to automatic processing. Heavy AI users moved from deliberate reasoning to pattern following behavior, which undermined their abilities to evaluate their understanding (Fernandes et al., 2024, p.14), meaning individuals begin to follow patterns and outputs from AI without the self reflection crucial to true understanding of content. This finding poses a challenge for educational settings, as it provides evidence for the fact that simply teaching students about AI usage may not improve their metacognitive accuracy, and could actually worsen the issue.

To accurately assess their AI assisted learning performance, one might assume that a better understanding of AI would be critical. This is not always the case, as proven by Fernandes' study. The researchers checked AI literacy using the SNAIL scale (Laupichler et al., 2023), which is often used to assess the AI literacy of non-experts. Here Fernandes found a paradoxical finding, where when participants had higher AI literacy, they also had less accurate self-assessment (Fernandes et al., 2024). In other words, knowing more about how AI works increased confidence whilst reducing accuracy for self-assessment. This relationship is supported by the correlation data, which illustrated a positive correlation between SNAIL understanding and the estimation and performance gap at $r = 0.21$, $p < .01$ (Fernandes et al., 2024). How this paradox works is that students who understand AI may be more inclined to be confident about their ability to use the tools, leading to a phenomenon that researchers call the "better than average effect," (Fernandes et al., 2024, p. 14) where familiarity with a subject causes baseless confidence.

Another key finding from the Fernandes study suggests that AI use appeared to eliminate the traditional Dunning Kruger effect. Normally, in non AI contexts, the Dunning-Kruger effect describes a cognitive bias where low performers overestimate their abilities, while high performers underestimate theirs. However, the Fernandes paper proved that AI leveled performance between high and low performers whilst doubling metacognitive bias for the entire sample (Fernandes et al., p. 14). This means that AI raises everyone's performance to similar levels while also making everyone overconfident, eliminating the relationship found in the Dunning-Kruger effect between skill level and metacognitive accuracy. This draws into consideration a

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

concerning implication. While the usage of AI reduces gaps between high and low performers, it creates a problem where all students, regardless of individual skill and ability, become equally poor at self-assessment. These findings suggest that AI assistance creates a different and potentially more problematic pattern of performance than traditional answer keys, where performance is improved while metacognitive monitoring at all learners' skill levels is diminished.

COMPARATIVE ANALYSIS

Sections 3 and 4 together reveal both similarities and differences that are important in how traditional answer keys and AI assistance affect the metacognitive accuracy of learners. The main similarity is that they both lead to fluency-based illusions, where the ease of obtaining answers misleads learners about their understanding of subjects. In both cases, students often misunderstand their fluency as proof of genuine learning, as proven by Wang's study finding that fluency affects metacognitive judgment. From here, the key point is that the two methods differ greatly in the magnitude and pattern of their effects. Traditional answer keys, especially when used with questions that cause students to reflect on their learning, could actually support accurate self-assessment, while AI assistance consistently undermines it. These differences suggest the source of correct answers matters as they actively shape students' performance and metacognitive processes in diverse ways.

Table 1

Comparison of Traditional Answer Keys and AI Assistance on Student Performance, Fluency Illusions, and Metacognitive Accuracy.

Dimensions	**Traditional Answer Keys**	**AI Assistance**
Effect on performance	Improves objective performance (Tomanek et al., 2017)	Improves objective performance (Fernandes et al., 2024)
Fluency based illusions	Creates fluency-based illusions (Carpenter et al., 2013; Wang & Xing, 2019)	Amplifies fluency based illusions due to the depth and comprehensiveness of AI responses (Fernandes et al., 2024)
Effects on Metacognitive Accuracy	Can support accurate self reflection when paired with self reflection prompts. (Tomanek et al., 2017)	Consistently erodes metacognitive accuracy. (Fernandes et al., 2024)
Dunning Kruger Effect	Maintains the traditional DKE pattern: Low performers overestimate, and high	Eliminates the DKE effect, as all skill levels become equally overconfident

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

	performers underestimate. (Fernandes et al., 2024; Noushad et al., 2023)	(Fernandes et al., 2024)
Processing Shift	Shorter answers naturally cause reflection. (Tomanek et al., 2017)	Causes learners to shift from reflective to automatic processing. (Fernandes et al., 2024)

Key difference 1: Traditional methods support metacognitive accuracy when paired with reflection

A critical finding from the Tomanek study illustrates that answer keys do not always create metacognitive problems, proving the importance of the design of the feedback system. Enhanced answer keys included not only correct answers, but explanatory information that encouraged students to reflect on their processes and understanding. The effectiveness of such modes of answer provision were highly effective, as proven by the fact that they helped "students to increase their understanding of the content and engage in... three dimensions of metacognition: intelligibility, plausibility, and wide applicability" (Tomanek et al., 2017, p. 11; Grotzer & Mittlefehldt, 2012). These reflection questions work by adding a delay between receiving answers and making metacognitive judgements, which counteracts the immediate fluency effects of non-enhanced answer keys. Additionally, it is important to note that students who used these enhanced answer keys claimed they helped to identify what they needed to continue studying and gained greater understanding of concepts they were learning. This demonstrates that traditional answer key provision can support accurate metacognition when it includes reflection that truly improves understanding rather than just answer checking.

Key difference 2: Traditional methods preserve the DKE, whilst AI eliminates it

The two methods also differ in how they affect the relationship between skill level and self-assessment. Specifically, they differ in the way they affect the Dunning-Kruger effect. Traditionally, low performers tend to overestimate their performance whilst high performers underestimate their performance, creating different challenges for metacognition at different skill levels. With traditional answer keys, the pattern generally stays the same. However, "the classic DKE, where lower performers overestimate and higher performers underestimate their performance, disappeared with AI use" (Fernandes et al., p. 14). This indicates that AI leads to similar overconfidence over all skill levels from low to high, removing the natural variation in metacognitive bias that normally exists. While AI unifies performance by helping all to be better, it also unifies metacognitive failure by making everyone equally bad at assessing their true abilities and understandings regardless of their skill level.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

Understanding why AI leads to more severe metacognitive problems requires looking at what makes AI assistance in education fundamentally different from traditional answer provision. The key difference lies in the great detail and authority that is provided with an AI's response that provides not only correct answers but also detailed and fluent explanations. According to Carpenter's findings on fluency, it generally increases perceived learning without actually improving learning, and AI's comprehensive and clear answers inevitably amplifies this effect far beyond the effect in which simple answer keys produce. Consequently, heavy reliance on AI will then produce a cognitive shift away from reflective and knowledge-based thinking to a more automatic one, which will ultimately weaken the learner's ability to critically self assess their understanding (Fernandes et al., p. 14). This could also be explained by the fact that when students receive comprehensive answers that answer their questions fully, they have less of a reason to think critically about the answers that would reveal gaps in their knowledge. This leads to a different learning environment than when traditional answer keys were used, where shorter answers naturally caused students to think more about the responses they received.

This comparative analysis reveals an important point that is often overlooked when discussing educational technology: The source of correct answers is an important variable that greatly shapes students' learning and understanding. Both traditional and AI assistance improve performance of learners by providing answers to questions, but they create different effects on metacognitive processes. Traditional methods, when enhanced with reflection, can maintain or even improve metacognitive accuracy while also supporting learning. On the other hand, AI produces patterns of improvement in performance often at the expense of metacognitive monitoring skills. This difference then means that it has great implications for educational design. This also means that educators must not only look at whether an educational tool improves learners' performance, but also the effects it has on students' awareness of their own understanding. Therefore there exists a gap between performance and metacognition that is caused by AI that requires different strategies for usage than those developed for traditional answer provision and learning tools.

IMPLICATIONS AND RESEARCH QUESTIONS

These findings are essential for how we design technology for education, particularly tools that provide answers to learners. First, answer keys should be further enhanced with reflection prompts that push learners to evaluate their understandings rather than just checking correctness to "increase their understanding of the content and engage in" (Tomanek et al., 2017, p.11). Additionally, AI interfaces should have built in mechanisms that counteract the immediate overconfidences that they produce, such as "Post-task reflection to encourage users to evaluate their performance after interacting with AI" (Fernandes et al., p. 15). The key takeaway from these two improvements is that educational tools should not solely be assessed for the extent of their performance enhancement. The designers should also optimize the tools for their effects on metacognition.

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

These implications then raise significant questions that future researchers should address: Could AI interfaces ever be designed to preserve their metacognitive accuracy, or is the tradeoff between the two inevitable?

LIMITATIONS AND FUTURE DIRECTIONS

While the current body of research clearly demonstrates a trend in how the source of an answer affects metacognition, there are still limitations to the research that future researchers must address. Most existing research examines single sessions of immediate testing without observing long term effects of repeated AI usage on learners' metacognition. Additionally, further research is required on domains beyond logical thinking and problem solving, such as STEM and creative writing. Similarly, to truly ascertain the effects of AI on metacognition, there must be studies with diverse student populations over different age groups to understand how results vary. Therefore, understanding individual differences, why some students are affected by fluency-based illusions more than others, could aid in personalized instruction that could further support metacognitive accuracy.

CONCLUSION

The performance metacognition gap created by AI assistance represents an important challenge that must be addressed when expanding educational technology beyond a single tool or platform. As AI becomes more common in classroom settings, understanding the metacognitive effects they have on students becomes increasingly important. This review demonstrates that there must be a shift in the way educational tools are being evaluated, moving from a performance based metric to a metric that includes both performance and metacognitive outcomes. Furthermore, the source of the correct answer is not a neutral variable, but rather an integral part of the learning process, as different sources lead to different learning outcomes with distinct consequences on metacognition. Effective learning does not just require success, but also a deep understanding of one's own understanding. Achieving this goal of performance improvement whilst maintaining or improving metacognitive accuracy requires careful attention to how the educational tools are designed and how they provide answers to learners.

REFERENCES

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). MIT Press.

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes (Vol. 2, pp. 35-67). Erlbaum.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), Psychology and the real world: Essays illustrating fundamental contributions to society (pp. 56–64). Worth Publishers.

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. Psychonomic Bulletin & Review, 20(6), 1350-1356.

Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2024). AI makes you smarter, but none the wiser: The disconnect between performance and metacognition. arXiv preprint arXiv:2409.16708.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34(10), 906-911.

Grotzer, T. A., & Mittlefehldt, S. (2012). The role of metacognition in students' understanding and transfer of explanatory structures in science. In A. Zohar & Y. J. Dori (Eds.), Metacognition in science education: Trends in current research (pp. 79-99). Springer.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. Journal of Experimental Psychology: General, 126(4), 349-370.

Koriat, A. (2010). Metacognition and mindreading: Judgments of learning for self and other during self-paced study. Consciousness and Cognition, 19(1), 251-264.

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. Developmental Science, 13(3), 441-453.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2), 187-194.

SCHOLARLY REVIEW

SCHOLARLY REVIEW JOURNAL

E-ISSN: 2996-8380
Published by Leadership & Innovation Lab

SPRING 2026

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? Psychological Science, 19(6), 585-592.

Laupichler, M. C., Aster, A., Haverkamp, N., & Raupach, T. (2023). Development of the "scale for the assessment of non-experts' AI literacy"—An exploratory factor analysis. Computers in Human Behavior Reports, 12, 100338.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 26, pp. 125-173). Academic Press.

Tomanek, D., Sabel, J. L., Schroeder, N. L., Simon, A., Schaertl, E., & Wichmann, A. (2017). Using enhanced answer keys and reflection questions to support students in learning biology. CBE—Life Sciences Education, 16(3), ar40.

Wang, J., & Xing, Q. (2019). Metacognitive illusion in category learning: Contributions of processing fluency and beliefs. Advances in Cognitive Psychology, 15(2), 100-110.

Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning? Learning and Instruction, 49, 199-217.