



Calibration Discrepancy Predicts Students' Subsequent Metacognitive Strategy Use in Computer-based Learning Environments

HaeJin Lee¹ · Nigel Bosch^{1,2}

Received: 15 November 2024 / Accepted: 3 September 2025
© The Author(s) 2025

Abstract

Students often misjudge their understanding of learning material, which can lead to the use of ineffective learning strategies and result in suboptimal learning outcomes. However, it remains unclear how misjudgments relate to the use of metacognitive strategies in online learning settings, which is essential context for developing effective interventions that support students in making (and using) accurate judgments of their performance. To address this, we analyze data from 210 college students using a computer-based learning environment, investigating the relationships among calibration discrepancy, judgments, and strategies, as well as the factors affecting shifts in metacognitive judgments during learning. Students who overestimated their pretest retrospective judgments engaged less in metacognitive strategies, particularly in preparatory actions before quizzes ($b = -9.100$, $p < .001$). Notably, pretest retrospective judgments—rather than actual pretest scores—significantly predicted students' engagement in these metacognitive strategies ($b = -9.841$, $p < .001$). Furthermore, increased engagement in repeated quiz-taking was a significant negative predictor of changes in metacognitive judgments ($b = -1.792$, $p = .036$), indicating that students engaging in repeated quizzes tended to adjust their judgments more conservatively. These results highlight the role of pretest retrospective judgments in shaping engagement with metacognitive strategies, underscoring the importance of correcting early calibration discrepancies. Our findings advocate for early, proactive metacognitive support tools that go beyond merely presenting information, offering guidance on interpreting feedback, and implementing strategies to better align students' judgments with their actual performance.

Keywords Metacognition · Metacognitive calibration interventions · Computer-based learning environments · Postsecondary education

Extended author information available on the last page of the article

Introduction

Metacognition, often described as thinking about one's thinking, encompasses essential cognitive activities such as planning, monitoring, reflecting, and evaluating one's learning strategies and knowledge (Dunlosky & Metcalfe, 2009; Flavell, 1979). The significance of metacognitive skills in enhancing learning outcomes is well-documented (Acosta-Gonzaga & Ramirez-Arellano, 2021; Dumford & Miller, 2018; Stanton et al., 2021; Van Der Stel & Veenman, 2014). Particularly in online learning environments—where students are expected to independently navigate their learning—accurately judging one's own knowledge and continuously adjusting learning strategies based on self-assessment becomes especially important (Broadbent & Poon, 2015; Kizilcec et al., 2017; Sorgenfrei & Smolnik, 2016).

A common metric used to assess how accurately one can judge their own knowledge or learning performance is known as *metacognitive calibration* (Alexander, 2013; Pieschl, 2009; Winne & Azevedo, 2014). In other words, metacognitive calibration refers to the degree to which students' self-assessments align with external assessments, such as test scores (Stone, 2000). Calibration is context-dependent and can be influenced by factors such as task difficulty (Boud et al., 2013; Nietfeld et al., 2005). Having accurate metacognitive calibration is crucial, since it enables students to allocate their time and effort appropriately; students can focus on learning new or challenging material while spending less time on topics they know they have already mastered (Chou et al., 2015). Such strategic allocation of time and study resources likely contributes to the positive correlation observed between metacognitive calibration and learning outcomes (Kelemen et al., 2007; Lin & Zabricky, 1998; Wiley et al., 2005).

Despite the significance of metacognitive calibration, students often struggle to accurately assess their own performance (Bol & Hacker, 2001; Chen, 2003). This calibration discrepancy, where students make inaccurate judgments of their knowledge, can lead to ineffective learning strategies and hinder academic progress (Klassen, 2002). The challenges students face in self-assessing their understanding are particularly pronounced in online learning environments (Bringula et al., 2021), where limited external feedback can exacerbate the issue (Chou & Zou, 2020). Online learners often miss out on real-time interactions with instructors and peers, which are crucial elements for evaluating performance. This lack of external reinforcement and guidance can widen the gap between students' perceived and actual performance, leading to overestimation or underestimation of their knowledge and skills (Bol et al., 2005). Such discrepancies may result in neglecting areas needing improvement or overstudying familiar concepts.

Overestimation of one's knowledge is especially concerning because students frequently exhibit overconfidence when assessing their performance (Helzer & Dunning, 2012; Saenz et al., 2017). Overconfidence often leads students to adopt ineffective learning strategies and reduce time and effort devoted to studying, resulting in sub-optimal learning outcomes (Aghababayan et al., 2017; Dunlosky & Rawson, 2012). Research has also identified a notable disparity in metacognitive calibration between low- and high-achieving students (Pennycook et al., 2017). Typically, low achievers exhibit more overconfidence—a phenomenon known as the Dunning–Kruger effect

(Kruger & Dunning, 1999). In contrast, high-achieving students often have better calibration (Harris et al., 2009), though they can sometimes exhibit underconfidence despite their superior actual performance (Hacker & Bol, 2004).

Given these concerns, numerous forms of interventions have emerged to improve students' metacognitive calibration (Bol & Hacker, 2001; Foster et al., 2017; Morphew, 2021; Nederhand et al., 2019), but the effectiveness of these interventions has been inconsistent (Saenz et al., 2019). For example, Emory and Luo (2022) discovered that metacognitive monitoring interventions did not enhance the calibration of community college students in online learning environments, underscoring the need for more targeted studies on the effectiveness of calibration training tools in online learning settings, especially for diverse student groups like those in community colleges. Moreover, tools supporting metacognitive calibration are predicated on the assumption that improved calibration leads to enhanced learning outcomes. Therefore, a typical strategy used in these interventions involves providing learners with feedback after assessments to highlight the differences between students' perceived performance and their actual performance, aiming to raise awareness of inaccuracies in their self-assessment skills (Foster et al., 2017; Hacker et al., 2000).

While recognizing the precision of one's judgment is crucial, the practice of providing students with feedback on performance discrepancies post-assessment primarily aims to correct superficial errors in their metacognitive evaluations. Further, this approach often fails to encourage students to reflect on and address the strategies that might have contributed to performance discrepancies. Dunlosky and Thiede (2013) further highlighted that providing accurate performance information alone (e.g., student-facing dashboards) does not automatically translate into deeper cognitive processing or improved learning strategies. Likewise, one of the fundamental gaps in performance information interventions is their inability to provide students with insights into why and how calibration discrepancy impacts students' subsequent engagement in learning behaviors, leading to potential suboptimal learning outcomes in online learning settings.

To overcome these limitations in post-assessment feedback regarding performance discrepancies from metacognitive interventions, it is necessary to uncover the relationships between calibration discrepancy (i.e., inaccurate judgments of one's performance), metacognitive judgments (i.e., one's perceived performance), and the subsequent engagement in metacognitive strategies. Understanding these relationships could support the development of calibration intervention tools which can go beyond mere identification of inaccuracies, guiding students to modify their learning strategies effectively to counterbalance and correct their pretest calibration discrepancies.

In this paper, we explore the relationship between pretest calibration discrepancy (i.e., at the beginning of a learning session) and the application of metacognitive skills in computer-based learning environments. We further look into how each initial actual and predicted grade relates to students' subsequent engagement in metacognitive strategies. Subsequently, we investigate how patterns of self-regulated learning (SRL)-relevant behaviors and other behavioral measures, such as the total number of topic transitions, serve as predictors for changes in students' metacognitive judgments.

Theoretical Models of Metacognition

Metacognition, defined as the awareness of one's cognitive processes (Flavell, 1979), is a higher-order thinking skill (Livingston, 2003). Metacognition encompasses a set of skills that enable learners to not only determine adequate learning strategies for various tasks but also continually assess their own knowledge and understanding (Dunlosky & Metcalfe, 2009). At its core, metacognition involves the regulation and monitoring of one's cognitive processes and awareness of one's own thought processes (Cheng & Chan, 2021; Nelson, 1990). Although cognition and metacognition are related, they differ in that cognition involves activities like information processing, problem-solving, and reasoning, whereas metacognition refers to the higher-order regulation and monitoring of these cognitive processes (Winne & Azevedo, 2014). Further, metacognition is a key component of SRL, which is known to be a critical element for academic success (Efklides, 2011; Panadero, 2017; Zimmerman & Moylan, 2009).

The concept of metacognition is multifaceted, with numerous models proposed for its understanding. Flavell's model of metacognition identifies four elements: metacognitive knowledge, experience, cognitive goals, and cognitive strategies (Flavell, 1979). Metacognitive knowledge, in particular, refers to an individual's understanding of personal, task, and strategy variables, encompassing self-awareness and insights about specific tasks and the strategies to approach them. Therefore, students with metacognitive knowledge are aware of their own beliefs, tasks, and strategies on how to approach a task. Metacognitive experience refers to the feelings that arise during the cognitive process. Flavell described that the four metacognitive elements—knowledge, experience, goals, and strategies—engage with each other in a joint process.

Brown (1987) defined metacognition as comprising two primary elements: cognitive knowledge and cognitive regulation. In this model, knowledge about cognition and regulation of cognition are interdependent, engaging in a recursive relationship. Cognitive regulation involves overseeing cognitive activities, including planning, monitoring, and evaluating. Schraw and Dennison (1994) echoed Brown's model by also subcategorizing metacognition into knowledge of cognition and regulation of cognition. However, Schraw and Dennison diverged in their detailed breakdown of knowledge of cognition, categorizing it into specific types: declarative, procedural, and conditional knowledge. Subsequent conceptualizations of metacognition, such as those proposed by Pintrich et al. (2000) and Efklides (2008), offer further insights to understand elements of metacognition. Pintrich et al. (2000) conceptualized metacognition involving three main components: (i) metacognitive knowledge, (ii) metacognitive judgment and monitoring, and (iii) self-regulation and control. In Pintrich's model, metacognitive judgment and monitoring correspond to one of the metacognitive elements in Flavell's model: specifically, metacognitive experience. Metacognitive judgments and monitoring, unlike the static nature of metacognitive knowledge, are dynamic and process-oriented, reflecting students' metacognitive awareness and their activities while performing tasks.

Efklides (2008) advanced a conceptualization of metacognition encompassing three key elements: metacognitive knowledge, experience, and skills. In Efklides's

model of metacognition, metacognitive knowledge encompasses information regarding persons, tasks, strategies and goals (Efklides, 2008). Metacognitive experiences, distinct from metacognitive knowledge, serve as a bridge between cognitive and emotional regulation. According to Efklides, metacognitive experiences manifest as metacognitive judgments/estimates, feelings, and task-specific knowledge in real-time (Efklides, 2001, 2006a). Therefore, when students make a prediction of their own knowledge (i.e., metacognitive judgment), they are engaging in metacognitive experience. On the other hand, metacognitive skills pertain to the intentional application of strategies for cognitive control. These metacognitive skills encompass a range of strategies, including planning, regulating cognitive processes, monitoring, and evaluating the results of task execution.

In the conceptual frameworks of metacognition developed by Flavell, Pintrich, and Efklides, metacognitive experience is identified as a separate and distinct element. Scholars agree that both metacognitive knowledge and experiences are two major elements of metacognition (Veenman et al., 2006). Although metacognitive experience is not identified as a separate main component in the models of Brown (1987), Schraw, and Dennison (Schraw & Dennison, 1994), it is considered as an essential skill within the metacognitive monitoring aspect of the metacognitive regulation component in their models (Zhao & Ye, 2020). Building on such insights, Efklides specifically highlighted the need for increased attention to metacognitive experience, pointing out the pivotal role of metacognitive experiences not only in metacognition but also in facilitating students' self-regulated learning (Efklides, 2006a, b, 2009).

Measuring Metacognitive Judgment, Metacognitive Skills, and Calibration Discrepancy

In online learning environments, accurately assessing the numerous dimensions of metacognition, including metacognitive judgment, skills, and calibration discrepancy, becomes critical for both understanding and enhancing learning. Comprehensive metacognition measurement is essential to grasp how students regulate and evaluate their learning processes in online learning environments. However, the precise classification of metacognitive calibration within the broader spectrum of metacognition—and thus its measurement—is not yet universally agreed upon (Akturk & Sahin, 2011). Despite this, researchers agree that calibration is a skill of metacognitive monitoring (Veenman et al., 2006), which refers to an individual's awareness of their cognitive processes, such as comprehension or task performance (Zhao & Ye, 2020). While metacognitive monitoring judgments refer to students' subjective evaluations (e.g., confidence ratings or judgments of learning), monitoring accuracy reflects how well those judgments correspond to actual performance, with calibration representing one specific measure of monitoring accuracy. From this viewpoint, calibration reflects the degree to which students' self-perceived performance aligns with their actual performance.

Therefore, metacognitive (mis)calibration has typically been measured by taking the difference between a student's perceived performance on an assessment and their actual performance (Emory & Luo, 2022; Pesout & Nietfeld, 2021). Additionally, measuring changes in metacognitive judgment is valuable for assessing how engage-

ment with educational content and feedback (i.e., results from taking the quiz) during the learning session might alter students' self-perceived understanding and mastery of the material. Understanding these changes can provide insights into how students' self-perceptions evolve through the learning process. Specifically, examining factors—such as engaging in specific learning patterns—associated with these changes can offer valuable insights into what could potentially shape shifts in students' metacognitive judgments.

Expanding from this focus on metacognitive calibration, the measurement of students' metacognitive strategies emerges as a similarly challenging task (Greene & Azevedo, 2010). Metacognitive strategies, a common facet of SRL, can manifest in numerous ways, involving a variety of learning strategies (Mitsea & Drigas, 2019). For instance, metacognitive strategies could encompass planning and monitoring one's learning strategies (Akturk & Sahin, 2011; Dunlosky & Thiede, 2013; Rivas et al., 2022). The complexity of measuring metacognitive strategies necessitates not only identifying various strategies but also understanding how they unfold over time (Azevedo, 2014). Researchers highlighted the importance of considering the temporal and sequential aspects when it comes to measuring SRL strategies (Molenaar & Järvelä, 2014). Particularly, incorporating temporality into understanding SRL emphasizes the dynamic nature of SRL, underscoring how SRL strategies evolve and adapt over time.

Consequently, the advent of computer-based learning environments enabled researchers to analyze the interaction log trace data generated by online learning platforms, providing a means to measure SRL events in real-time. Real-time data present a valuable opportunity to assess the use of metacognitive strategies, uncovering patterns of learning behavior. Thus, researchers adopted numerous temporal measures of SRL strategies in computer-based learning environments (Saint et al., 2022). Some data-driven approaches include process mining (Bogarín et al., 2018; Sobocinski et al., 2017), epistemic network analysis (Paquette et al., 2021), lag-sequential analyses (Kuvalja et al., 2014), and constrained sequential pattern discovery (Wong et al., 2019; Liu & Moon, 2023). Moreover, there exist approaches such as coherence analysis (CA) (Segedy et al., 2015) which provides a more theory-driven approach to measuring and understanding students' use of metacognitive strategies. CA is a method that evaluates metacognitive strategies by examining the sequential alignment of students' learning actions within online learning contexts (Segedy et al., 2015). In particular, CA examines how learning activities are sequentially connected in a way that indicates metacognitive strategies such as self-reflection and planning, shedding light on how students apply metacognition to navigate their learning processes.

Coherent actions do not have to occur in sequence, yet it is crucial to limit the time gap between the learning activities to be considered as coherent. In the context of Betty's Brain, prior research found that students generally employed information within five minutes of encountering it. Specifically, Segedy et al. (2015) found that coherent actions within five-minute timeframe showed a positive correlation with assessment scores during an online learning session and overall learning improvements throughout the entire learning session. Measuring metacognitive strategies via

CA can be adapted to quantify various facets of students' use of metacognitive strategies, tailored to specific learning settings and research contexts.

Several studies leveraged CA to measure students' use of metacognitive strategies in computer-based learning environments. For instance, Bosch et al. (2021) investigated the connections between verbalized metacognition and learning, confusion, and metacognitive problem-solving strategies. Zhang et al. (2020) initially employed CA in a computer-based learning environment called Betty's Brain to examine the relationship between confusion and metacognitive strategies. Building on this, their subsequent study Zhang et al. (2022) further utilized CA to explore the evolution of metacognitive strategy use, advancing the understanding of how metacognitive strategy use develops over time. Through a blend of data-driven and theory-driven approaches, researchers developed a multifaceted understanding of how students employ metacognitive strategies within computer-based learning environments. This comprehensive approach to measurement is crucial for accurately capturing and supporting students' metacognitive strategy use, ultimately enhancing their learning outcomes.

In this study, we use CA to measure students' use of metacognitive strategies by analyzing the temporal and sequential alignment of their learning actions captured through trace data. Although various approaches exist for assessing metacognitive strategy use (Azevedo, 2015; Veenman & van Cleef, 2019; Winne & Perry, 2000), CA is particularly well-suited for our context. CA has been employed in prior research to capture metacognitive behaviors in short-term, online learning environments (Bosch et al., 2021; Segedy et al., 2015; Zhang et al., 2022)—settings similar to ours, given the constrained scope and duration of the learning tasks in our study.

Interventions Supporting Metacognitive Calibration

Numerous studies have examined the relationships between various facets of metacognition and academic performance (Abdelrahman, 2020; Aloqleh & Teh, 2019; Avargil et al., 2018; Coutinho, 2007; Jalili et al., 2018; Narang & Saini, 2013; Pradhan & Das, 2021; Taraban et al., 2000; Wagener, 2016; Ward & Butler, 2019). Additional research measured the connections between metacognition and numerous learning strategies. For instance, Zhou (2023) discovered a moderate positive correlation ($r=.42$) between university students' metacognitive calibration and their online information search performance scores. Kubik et al. (2022) found that judgments of learning with incomplete information, such as presenting partial cues or word stems, can stimulate covert retrieval attempts, where individuals internally try to recall information without manifesting it externally (e.g., through speech or writing). Miller and Geraci (2011) explored the relationship between students' metacognitive judgment (i.e., exam predictions) and their confidence and found that low-performing students exhibited lower confidence compared to their high-performing peers. Moreover, Zhao and Ye (2020) investigated the impact of metacognitive calibration accuracy on assignments in online learning environment, finding that students with better calibration performed more effectively on both assignments and exams, although the time spent on assignments did not directly influence their performance. Sun et al. (2021)

found positive correlations between students' metacognitive experience and their test scores for English as a foreign language.

In computer-based online learning settings, the positive relationship between metacognitive strategies and academic performance underscores the importance of metacognitive strategies for successful learning (Cho & Heron, 2015; Goradia & Bugaric, 2017; Tsai et al., 2018). However, the inherent challenge in computer-based environments stems from the requirement for learners to demonstrate a higher degree of metacognitive skills and independence, which places a significant demand on students' self-regulatory abilities. While it is common for students to face challenges in accurately evaluating their own performance, calibration discrepancy is often not an indication of incompetence but rather a result of insufficient support or practice. However, evidence suggests that with appropriate support and guidance, students can improve their metacognitive calibration abilities (Abdelrahman, 2020; Urban & Urban, 2019). Consequently, a range of interventions have been developed to assist students in better self-evaluating and calibrating their knowledge both in online and traditional classroom environments (Foster et al., 2017; Hacker et al., 2000; Nietfeld et al., 2006; Saenz et al., 2017; Urban & Urban, 2019). However, the effectiveness of interventions has been inconsistent, reflecting the complexity and diversity of challenges encountered in educational contexts, including in online settings.

Among metacognitive calibration interventions, one common approach is providing feedback to students on the discrepancy between their expected grades and their actual performance on assessments. While feedback-oriented interventions might be beneficial for students to correct their calibration discrepancies in some cases, with only the information on their actual assessment score students may struggle to make meaningful adjustments. For instance, Foster et al. (2017) found that students' calibration accuracy did not improve upon receiving feedback on their estimated and actual test performance. On the other hand, Urban and Urban (2019) found that students showed improvements in calibration after receiving a combination of approaches (i.e., SRL training, peer evaluation and calibration feedback). Further, Saenz et al. (2017) compared the effectiveness of five different calibration-supporting interventions: (i) review (reviewing test questions and performance grade predictions), (ii) salient feedback (which entailed clear feedback on performance and prediction accuracy), (iii) motivation warning lecture (an educational session about personal motivations and academic information's role in prediction accuracy), (iv) incentives (students were given the chance to receive \$50 for making accurate predictions), and (v) reflective practices (students had to reflect on their predictions for an extended period receiving no feedback). Saenz et al. (2017) discovered that only the motivation warning lecture and salient feedback improved students' prediction accuracies. Likewise, there is a need for more targeted studies on the development of tools that effectively support metacognitive skills, particularly in online learning environments (Azevedo et al., 2022). There is significant potential for these calibration interventions to become more comprehensive, yet further research is needed to explore the interconnections among various facets of metacognition. For instance, insights into how students adjust their learning strategies based on their metacognitive judgments could enrich current feedback-oriented interventions by providing more targeted guidance on engagement in effective learning strategies.

Overview of the Study

Despite significant progress in understanding metacognition in learning (Tsai et al., 2018; Zhou, 2023), there remains much to be explored regarding how key metacognitive elements—calibration, judgment, and strategies—interact within computer-based learning environments. Prior studies showed that students whose metacognitive calibration is relatively accurate tend to achieve higher learning outcomes (Dunlosky & Rawson, 2012; Hacker et al., 2000; Nietfeld et al., 2005). Although only a few studies explored why calibration accuracy translates into better academic performance, existing evidence suggests that students who overestimate their knowledge often fail to recognize when self-regulatory strategies are needed, tend to ignore feedback, and are less likely to take corrective actions in goal-setting practices (Hadwin & Webster, 2013; Hattie, 2013). Although a limited number of semester-long studies linked calibration inaccuracy to later strategy use (e.g., Hadwin & Webster, 2013, for goal-setting), little empirical evidence addresses whether the same relationship holds in shorter, computer-based learning environments, which are becoming an increasingly common mode of instruction.

Additionally, high-performing students tend to engage more frequently in metacognitive strategies such as self-testing, planning, and monitoring (Miller & Geraci, 2011). These students adapt and adjust their learning strategies during learning, which may partly explain their frequent use of metacognitive strategies. However, these students often underestimate their performance—a pattern related to the Dunning-Kruger effect—suggesting that they are more conservative in evaluating their own knowledge. This conservatism may further prompt high-performing students to engage in strategies that support continuous self-assessment and regulation. Students' perceived performance may serve as an immediate driver of metacognitive strategy use. While prior studies have shown that what students believe they know directly influences their study choices (Metcalfe & Finn, 2008), these findings were primarily drawn from in-person settings involving college students. It remains underexplored whether similar patterns hold in computer-based learning environments. Investigating this question is particularly important in digital learning contexts, where early detection of learners' confidence levels could enable educational systems to deliver timely, targeted interventions that foster effective metacognitive strategy use and, ultimately, enhance learning outcomes.

Moreover, research shows that opportunities for self-evaluation can improve metacognitive calibration over time (e.g., Osterhage et al., 2019; Boud et al., 2013), and increased self-assessment may even lead to underconfidence in students' judgments (Koriat et al., 2002). However, most of these findings come from semester-long, in-person courses. Therefore, additional empirical evidence is needed to establish whether these patterns extend to computer-based learning environments. Zhao and Ye (2020) emphasized that while numerous studies focused on the role of metacognitive calibration in predicting student learning outcomes, the majority were conducted in controlled laboratory environments with simulated tasks (Thiede et al., 2003). Additionally, Zhao and Ye (2020) pointed out a significant gap in research exploring metacognitive calibration in natural learning settings, especially in online learning environments at the university level.

Motivated by these gaps in literature, we focus on three research questions: the first two examine the relationship between students' calibration discrepancy and judgments, and their application of metacognitive strategies in subsequent learning sessions. The last research question investigates whether SRL patterns and other behavioral measures can predict changes in metacognitive judgments. RQ1 investigates the relationship between students' pretest calibration discrepancies—defined as the difference between their predicted (guessed pretest) scores and actual pretest scores—and their subsequent engagement in using metacognitive strategies. This question seeks to understand how students' overconfidence or underconfidence, as evidenced by this difference, predicts their approach to learning. While RQ1 offers insights into the impact of calibration discrepancy on subsequent metacognitive strategy use, it does not dissect the individual components of calibration discrepancy—namely, students' pretest retrospective judgment and their actual performance. This limitation prompts our second research question, which aims to unpack these elements separately.

Therefore, RQ2 examines the individual components of calibration discrepancy, focusing on the independent relationships between students' pretest retrospective judgment, their actual performance, and their usage of metacognitive strategies. Lastly, in RQ3, we explore how certain SRL patterns and behavioral measures predict changes in students' metacognitive judgments, thereby providing insights into the specific factors influencing this shift in metacognitive judgments.

In sum, we answer the following research questions:

- **RQ1.** How do students' pretest calibration discrepancies relate to their use of metacognitive strategies?
H1. We hypothesize that higher levels of pretest calibration discrepancy will be associated with less frequent use of metacognitive strategies. Specifically, overconfident students are less likely to engage in reflective or regulatory learning behaviors, consistent with prior studies showing that such students often fail to recognize when self-regulatory strategies are needed, tend to ignore feedback, and are less likely to take corrective actions (Hadwin & Webster, 2013; Hattie, 2013).
- **RQ2.** How are students' pretest retrospective judgments and actual performances separately related to their metacognitive strategy use?
H2. We hypothesize that students' actual performance will be positively associated with metacognitive strategy use, whereas pretest retrospective judgments will not show a significant relationship. This hypothesis is based on prior studies which showed that high-performing students tend to engage more frequently in metacognitive strategy use (Miller & Geraci, 2011; Zimmerman, 1989).
- **RQ3.** How do SRL-relevant learning patterns, number of topic changes, and number of activity measures predict changes in students' metacognitive judgments?
H3. We hypothesize that SRL-relevant learning patterns involving self-assessment (e.g., taking a quiz in this study), along with the number of activity measures, will be negative predictors of changes in students' metacognitive judgments. That is, students who frequently engage in self-evaluations and demonstrate fre-

quent engagement with learning activities are expected to show a decrease in their perceptions of how much they learned, in line with prior work (Koriat et al., 2002). This may be because such behaviors provide opportunities for students to assess their understanding, increasing their awareness of knowledge gaps and leading them to make more conservative predictions of their performance.

Method

Participants

We collected behavioral trace data and survey responses from a total of 210 college students who learned four distinct subtopics of statistics through a web-based learning environment that we developed. Our sampling approach was twofold: first, we recruited 112 students from a large, public research university located in the Midwest of the United States. These students, who were enrolled locally, were given course credit for their participation in the study. Second, we used Prolific, a digital crowdsourcing platform that provides access to a pool of students from various U.S. colleges and universities (Peer et al., 2021). Prolific offers the option to filter participants based on several criteria, including demographic factors. For our study, we limited our participant criteria to include only undergraduate students from either two-year or four-year institutions, including both community colleges and universities. This second group comprised 98 students, each of whom received \$15 in compensation.

We provide the self-reported demographic data of our participants to describe the diversity within the sample, although demographic variables were not included in our analysis. The sample characteristics are fundamental in establishing the extent to which findings from meta-analytic research, which may include studies like this one, can be generalized. The local sample ($n = 112$) was composed of 56.3% White, 25.0% Asian, 8.0% Black, 8.0% Latinx/Hispanic, and 2.7% grouped for anonymity; 74.1% female, 23.2% male, and 2.7% other genders grouped for anonymity. The Prolific sample ($n = 98$) consisted of 55.2% White, 16.3% Asian, 12.2% Black, 11.2% Latinx/Hispanic, and 5.1% grouped for anonymity; 44.9% female, 44.9% male, and 10.2% other genders grouped for anonymity. The Prolific sample represented students from 62 unique colleges/universities, including 11 community colleges.

Online Learning System Settings

We used a self-guided, web-based online learning system that we developed in a prior study (Lee & Bosch, 2024), which allowed students to navigate learning content at their own pace. The source code for the system is publicly available at https://osf.io/j9h74/?view_only=a93f7b3649414b288933cc73fb188795. The system was optimized for desktop or laptop use, and participants were instructed to use these devices. This system recorded students' trace data at the activity level, capturing each interaction (e.g., navigation events, quiz answers) with associated timestamps. The self-paced online learning system included four distinct, illustratively presented sub-

topics with associated icons (Fig. 1). No formal navigation training was provided, as findings from our in-person think-aloud usability study ($N=5$) indicated that all students were able to navigate the system without difficulty. Specifically, participants were able to access each learning activity by clicking labeled icons, and the interface did not include any complex features that required additional training.

Each subtopic module included one reading, quiz, set of worked examples, and summary. Participants received instructions indicating that they would study introductory statistics for 60 min by freely engaging with four types of learning activities before beginning the study. Students were not required to complete studying all subtopics during the learning session, nor all parts of each subtopic, as the system was open-ended. Additionally, the system allowed students to revisit and complete any activity multiple times, catering to their individual learning needs and preferences. Each module comprised four distinct learning activities: reading, quizzes, examples, and summaries.

All learning content used in the four learning activities was adapted from two main sources: (1) *Statistics: The Art and Science of Learning from Data*, 3rd edition, by Alan Agresti and Christine Franklin (Pearson Education, 2013), and (2) publicly available instructional content from the Pennsylvania State University's STAT 200 course. The content from both sources was developed for an introductory-level statistics audience and was adapted to match the readability level appropriate for college students without prior statistics experience—the target population of our study. All materials consisted primarily of text with static images and diagrams, some of which were adapted from the same textbook by Agresti and Franklin.

Students had the flexibility to decide the order of learning activities that they wanted to complete, irrespective of the subtopic. Every activity served a distinct learning purpose. The reading activity, typically four to six pages per subtopic, pro-

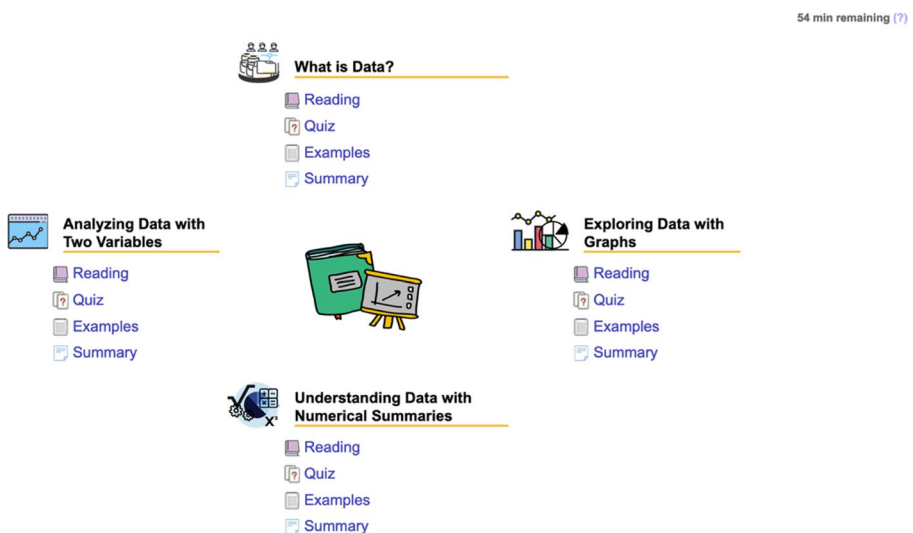


Fig. 1 Screenshot showing the learning software's main menu (top) alongside an illustration of a student's attempt at an incorrect quiz question (bottom)

vided comprehensive information about the subject matter. The quiz, consisting of around 10 questions, allowed students to assess their understanding of the material by taking it as many times as they wished. Incorrect answers were flagged, but the correct answers were not revealed to promote students' self-guided learning. The examples provided more than just correct answers to example questions; they demonstrated the proper problem-solving methods. Finally, the summary provided a concise recap of each module's essential learning materials, allowing students to review each subtopic's content quickly.

Study Procedure

The study required students to complete surveys and participate in a focused 60-minute learning session on four subtopics in introductory statistics. Before participation, students filled out a consent form (approved, along with all other study procedures, by our institutional review board). Students then completed a demographic survey, collecting information on race, gender, first-generation status, and other sensitive attributes. After completing the survey, students took a pretest consisting of 12 required multiple-choice questions designed to assess their prior knowledge of the topics covered in the subsequent learning session. Following the pretest, students were asked to estimate their score (0–100%) to reflect on what they thought they had achieved on the pretest, which was used to measure their pretest retrospective judgments. Students were not informed of their actual pretest scores after making these estimates. Students then began a self-paced learning session to study at their own pace, during which a timer was visible to show the remaining time, which only counted down when there was active interaction with the system, to promote focus and engagement. After the learning session, students took a posttest on the material they had studied. As with the pretest, students were asked to estimate their posttest score (0–100%), reflecting on what they thought they had achieved on the posttest to measure their posttest retrospective judgments.

Pretests and posttests were designed to be as similar as possible in both difficulty and subtopic coverage. We created two interchangeable versions of the test, labeled A and B, which could be used as either pretests or posttests. Although we intended the tests to be identical in difficulty and content coverage, we also implemented a counterbalanced test ordering to account for any potential differences. Students were randomly assigned one of two test orders: those in version A took test A as the pretest and test B as the posttest, while students in version B had the order reversed.

The test items reflected a mix of recall, comprehension, and application. Questions involving calculations (e.g., calculating means or medians) assessed application-level skills, while items focused on conceptual definitions in introductory statistics assessed recall and comprehension. Both the pretest and posttest consisted of 12 multiple-choice questions, with three questions aligned to each of the four subtopics to ensure balanced content coverage. To establish content validity for the pretest and posttest, four content experts (i.e., individuals with substantial post-graduate training in statistics) were asked to match randomly shuffled questions from Tests A and B, which were used interchangeably as pretests and posttests, based on the statistical concepts each question assessed. All four experts achieved 100% accuracy in this

matching task, indicating that the two test forms measured equivalent knowledge. We direct readers to our prior study (Lee & Bosch, 2024), which provides a detailed explanation of how the pretest and posttest were comparable in both difficulty and subtopic representation, along with the full set of test questions.

Data and Variable Measurements

The variables in our study were derived from students' behavioral trace data, recorded by the online learning system. Some variables directly reflected students' responses, such as metacognitive judgments, while others, related to student learning behavior—such as the use of metacognitive and SRL strategies—were measured using CA. In this context, CA was operationalized by analyzing the sequential alignment of students' learning activities. We identified two coherent activities (i.e., coherent quiz and reading) that signify specific metacognitive strategies. Further details on these measurement methods are provided in this section.

Metacognitive Judgment, Calibration Discrepancy, and Changes in Metacognitive Judgments

We constructed three metacognitive metrics: metacognitive judgment, calibration discrepancy, and change in metacognitive judgment. Pretest (i.e., before the learning session) and posttest (i.e., after the learning session) metacognitive judgment represents students' perceived understanding or beliefs about their own knowledge of the material before engaging in the learning session. This metric was measured by asking students to estimate their pretest and posttest scores as percentages, within a range of 0 to 100% (i.e., "What do you think your grade will be on the test you just took? (0–100%)"). Pretest calibration discrepancy quantifies the discrepancy between students' perceived performance (pretest retrospective judgment) and their actual performance on the pretest. To measure students' pretest calibration discrepancy, we subtracted students' actual pretest scores from their estimated scores (i.e., pretest retrospective judgments), using this difference as a measure for pretest calibration discrepancy. This discrepancy metric allows us to understand the variance between the students' pretest retrospective judgments and their actual test outcomes. A positive calibration discrepancy indicates overconfidence, i.e., that students estimated a higher grade than they actually achieved. For example, if a student estimated a pretest grade of 50% but actually scored 40%, the pretest calibration discrepancy would be +10%, indicating overconfidence in their knowledge. Conversely, a negative value signifies underconfidence, where students underestimated their performance.

Change in metacognitive judgment measures the evolution of students' self-assessments from before to after the learning session. We measured students' change in metacognitive judgments by taking the difference between posttest and pretest retrospective judgments. A positive change in metacognitive judgment indicates that a student's posttest retrospective judgment was higher than the pretest retrospective judgment, meaning students adjusted their self-assessments upwards. This upward adjustment possibly reflects increased confidence or perceived learning gains after the learning session. Conversely, a negative change in metacognitive judgment indicates

that students' pretest retrospective judgment was higher than their final estimate. This downward adjustment suggests that students revised their self-assessments downwards, potentially reflecting a decrease in confidence or a more accurate understanding of their knowledge following the learning session.

- **Pretest and posttest retrospective judgment:** Students' estimated pretest or posttest scores. These measures reflect students' perceptions or beliefs about their knowledge of the material before or after the learning session.
- **Pretest calibration discrepancy:** Pretest calibration discrepancy is calculated as follows: *estimated pretest grade* – *actual pretest grade*. This metric quantifies the accuracy of students' pretest retrospective judgments, where the values indicate overconfidence or underconfidence in their understanding.
- **Change in metacognitive judgment:** This measurement was calculated as *post-test retrospective judgment* – *pretest retrospective judgment*, measuring the shift in students' self-evaluations post-learning session, highlighting changes in confidence or perceived learning gains.

Use of Metacognitive Strategy

We measured two types of metacognitive strategies via CA: *coherent reading* and *coherent quiz*. Coherent reading and coherent quiz differ by their focus and timing in the learning process. Coherent reading activity refers to learning sequences in which students, after completing a quiz, engage with additional materials to address potential knowledge gaps. These activities are considered coherent when the subsequent learning action occurs within a five-minute window (Segedy et al., 2015), indicating a timely and purposeful response to taking a quiz. For instance, the sequence *Quiz* → *Read* reflects a proactive effort by students to use quiz feedback to locate and review relevant material, suggesting intentional efforts to improve understanding. Similarly, *Quiz* → *Examples* represents students reviewing worked-out examples after a quiz, which implies a coherent attempt to seek help and clarify misconceptions, particularly on missed items. In the *Quiz* → *Summary* sequence, students turn to summary content following a quiz, indicating an effort to consolidate key concepts and reinforce comprehension based on quiz performance. These sequences exemplify students' metacognitive strategies to monitor and regulate their learning in response to assessment outcomes, thereby qualifying as coherent reading activities. As such, coherent reading is one type of metacognitive strategy applied following a quiz. This coherent approach requires students to detect their knowledge gaps revealed by the quiz results and promptly direct their attention to bridging these gaps.

Time was tracked at both the activity and page levels using the trace data, which recorded every interaction (e.g., navigation events, quiz answers) with associated timestamps. This granularity allowed us to measure students' engagement with specific content related to each quiz item. For example, if a student answered quiz questions incorrectly on a subtopic like "What is Data", particularly on sample vs. population, we tracked the time spent reviewing the relevant pages in the Reading, Summary, and Examples activities.

Coherent quiz behavior, on the other hand, centers on the time *before* taking a quiz, involving proactive engagement with reading materials, examples, and summaries to prepare, reflecting a strategic approach in acquiring and assessing knowledge. While coherent reading is a response to identified weaknesses, coherent quiz represents preparatory actions for knowledge evaluation. Hence, coherent quiz behavior demonstrates that students are intentionally spending time on reading and understanding relevant information prior to evaluating their knowledge by attempting the quiz. For instance, if a student goes through the summary pages of one subtopic and takes the quiz (i.e., *Summary* $\rightarrow$ *Quiz*) for that subtopic within a five-minute window, then the student's activity sequence is considered a coherent quiz activity.

We measured coherent reading by adding up the time students devoted to reviewing material related to the questions they missed in the quiz, within a five-minute period after the quiz. Such a strategy demonstrates a student's proficiency in monitoring their academic advancement, discerning their own errors, and executing actions for improvement, which are fundamental elements of employing metacognitive strategies. We measured coherent quiz behavior by calculating the total amount of time students devoted to reading activities in the five-minute window before attempting quizzes related to those reading topics. In the context of this study, reading activities encompass three distinct types, which are studying the main reading content, going through worked-out examples, and going over the summary pages. If a student takes a quiz on one subtopic and then reviews the reading pages related to the questions they answered incorrectly within a five-minute window, this activity is considered a coherent reading activity.

SRL and Other Behavioral Measures

We measured students' use of SRL strategies by examining the sequence of learning activities (i.e., reading, quizzes, examples, and summaries) that students engaged in during learning (Lee & Bosch, 2024). In the previous study, we used sequence mining, especially constrained Sequential Pattern Discovery (cSPADE), which allowed us to identify frequent learning sequences of activities (e.g., Read $\rightarrow$ Quiz) and associate these sequences with potential SRL strategies (e.g., Read $\rightarrow$ Quiz with *seeking evaluation*), as detailed in Table 1. These associations were grounded in the literature (Corrin et al., 2017; Sonnenberg & Bannert, 2015; Zimmerman & Pons, 1986), and we mainly adopted Zimmerman's 14 classes of SRL strategies as a framework to relating learning patterns to potential SRL strategies (Table 1). Zimmerman and Pons's 14 SRL strategies (Zimmerman & Pons, 1986), which define SRL as observable actions directed at acquiring knowledge or skills. This framework aligns with our focus on students' active learning behaviors during the performance phase, as captured through trace data in a computer-based learning environment. For an elaboration on how each association between learning pattern and SRL-relevant strategies, see Lee and Bosch (2024).

For example, the sequence Read $\rightarrow$ Quiz indicates that students first engaged in a reading activity and then completed a quiz. This sequence suggests the use of the seeking evaluation strategy, as it reflects students' efforts to assess their understanding by reading the material and then taking the quiz. Using this SRL strategy frame-

Table 1 Learning patterns and associated SRL-relevant strategies with descriptions

Learning pattern	SRL-relevant strategy	Description	Support
Read → Quiz	- Seeking evaluation (Zimmerman & Pons, 1986)	When students read material and then take a quiz on it, they are evaluating their understanding and recall of the material they just read. After taking the quiz, they can gauge the quality or progress of their work based on their performance.	0.850 (<i>n</i> = 179)
Quiz → Read	- Keeping records and monitoring (Zimmerman & Pons, 1986) - Seeking information (Zimmerman & Pons, 1986) - Search (Sonnenberg & Bannert, 2015)	Students taking the quiz and then reading the main material signifies that students are aware of the knowledge gap and might specifically look for information to address the gaps.	0.770 (<i>n</i> = 162)
Quiz → Quiz	- Rehearsing and memorizing (Zimmerman & Pons, 1986) - Repeating (Sonnenberg & Bannert, 2015)	When students encounter the first quiz, they are prompted to recall specific information. By the second quiz, they are not just accessing their foundational understanding but also relying on memory from the previous quiz attempt.	0.690 (<i>n</i> = 145)
Quiz → Examples	- Keeping records and monitoring (Corrin et al., 2017) - Seeking information (Zimmerman & Pons, 1986) - Help-seeking (Corrin et al., 2017)	After taking the quiz, students are trying to make an effort to gather specific, detailed information on how to approach or solve problems correctly.	0.630 (<i>n</i> = 133)
Read → Examples	- Seeking information (Zimmerman & Pons, 1986) - Elaboration (Sonnenberg & Bannert, 2015)	Illustrates student-initiated efforts to seek additional knowledge from additional resources to bolster their learning. Students are actively seeking clarity and deeper understanding as Examples provides them with detailed worked-out problems with explanations on how to approach solving the problem.	0.570 (<i>n</i> = 120)
Quiz → Summary	- Keeping records and monitoring (Zimmerman & Pons, 1986) - Seeking information (Zimmerman & Pons, 1986) - Search (Sonnenberg & Bannert, 2015)	Students are not only keeping records of their quiz performance but are also actively seeking to enhance their understanding through the supplementary information provided in the summary. This dual approach allows them to both identify areas of improvement from their quiz results and address those areas by going through a summary.	0.450 (<i>n</i> = 95)

work established in the previous study (Table 1), we applied six different learning patterns and their associated SRL-relevant strategies as SRL constructs in this study. We counted the occurrences of each learning pattern to measure students' use of SRL skills.

Note. The leftmost column displays the frequent learning sequences, while the subsequent column lists the corresponding type of SRL strategy that each frequent learning sequence exemplifies. The last column indicates the proportion of students

who engaged in each frequent learning pattern at least once, as well as the corresponding number of students.

Alongside the SRL strategy measurements, we incorporated two additional behavioral metrics: the total number of topic transitions and the aggregate count of learning activities. We quantified the number of topic transitions by tracking the counts of students shifting between different topics during their learning sessions. Within a session containing four unique subtopics, we tallied the number of times students moved from one subtopic to another. Additionally, we calculated the total number of activities (i.e., reading, quizzes, examples, and summaries) in which students participated throughout their study. We were interested in these two measurements because a high or low total number of topic transitions and learning activities could potentially predict students shifting their metacognitive judgments.

Results

For all regressions in **RQ1-RQ3**, we checked the assumptions of linear regression, including linearity, independence, homoscedasticity, and normality, using the visualizations (i.e., residual plots, Q-Q plots), and found no significant deviations. For **RQ2** and **RQ3**, there was no evidence of multicollinearity; correlations and variance inflation factors are presented in Appendix Tables 5 and 6. We also conducted sensitivity power analysis using G*Power (Faul et al., 2009). With a sample size of $N=210$, an alpha level of 0.05, and a power level of 0.90, the minimum effect sizes were $f^2 = 0.051$ for **RQ1**, $f^2 = 0.061$ for **RQ2**, and $f^2 = 0.094$ for **RQ3**, each corresponding to a small-to-medium effect size (Cohen, 1988).

Descriptive Statistics

Table 2 summarizes the means and standard deviations of participants' performance, pretest and posttest retrospective judgments, and calibration discrepancy. Students demonstrated a significant learning gain, calculated as the difference between posttest performance and pretest performance ($d=0.705$, $t(209)=10.117$, $p<.001$), indicating a substantial improvement in knowledge after the learning session. Students studied an average of 3.952 out of 4 subtopics, with a subtopic counted as studied if they engaged in at least one of the four learning activities (i.e., Reading, Quiz, Examples, or Summary), suggesting that most students engaged in at least one activity for each subtopic. On average, students' pretest retrospective judgments were lower than their final judgments, suggesting that after the learning session, students estimated higher scores on their assessments (mean pretest retrospective judgment: 55.7%, mean posttest retrospective judgment: 73.4%). Both pretest and posttest calibration discrepancy measures (pretest: 0.796% points; posttest: 3.89% points) suggest that students demonstrated overall accurate calibration, which may be explained by established moderators of calibration accuracy, such as the increased accuracy of delayed versus immediate judgments and the effect of pretesting (Nelson & Dunlosky, 1991). While we could not directly analyze these potential moderators of calibration accuracy in

Table 2 Means and standard deviations of Performance, metacognitive Judgments, calibration Discrepancy, and coherent actions.

Category	Mean	SD
Initial performance (0–100%)	54.9%	19.0%
Pretest retrospective judgment (0–100%)	55.7%	20.2%
Pretest calibration discrepancy (percentage points)	0.796	20.6
Final performance (0–100%)	69.5%	21.6%
Posttest retrospective judgment (0–100%)	73.4%	17.4%
Posttest calibration discrepancy (percentage points)	3.89	17.6
Coherent reading (in minutes)	1.8	2.7
Coherent quiz (in minutes)	11.2	11.0

The measures provide an overview of students' performance and metacognitive judgments at both the initial stage (i.e., before the learning session) and the final stage (i.e., after the learning session), as well as their calibration discrepancy. The metrics on coherent reading and quiz show students' time engagement in these activities, measured in minutes. All statistics are based on data from $N=210$ students.

this study, given the fixed study design, these factors are discussed in the Discussion section as possible explanations for the observed calibration results.

Moreover, students' engagement in coherent quiz activity far exceeded their engagement in coherent reading (mean coherent reading: 1.8 min; mean coherent quiz: 11.2 min). This difference likely arises from how coherent reading and quiz activities were measured, as coherent reading only accounts for the time spent reviewing material students answered incorrectly on the quiz. Consequently, students who performed well on the quiz had fewer opportunities for coherent reading compared to students who scored lower. We examined the associations between students' engagement in coherent activities and their posttest performance. The results showed a significant positive association between coherent quiz engagement and posttest grade ($\rho=0.381$, $p<.001$), and a weaker but still significant association for coherent reading engagement ($\rho=0.174$, $p=.012$), suggesting that students who engaged more frequently in coherent activities tended to achieve higher posttest scores.

RQ1. How do students' pretest calibration discrepancies relate to their use of metacognitive strategies?

We conducted two linear regression analyses to address RQ1, which examines whether students' pretest calibration discrepancy serves as a predictor of their subsequent use of metacognitive strategies (i.e., coherent reading and quiz activities) during the learning session. In the first model, the dependent variable was the engagement in coherent quiz activities, with students' pretest calibration discrepancy as the independent variable. The second model was analogous to the first, with coherent reading engagement as the dependent variable. Our analysis revealed that students' pretest calibration discrepancy was a significant negative predictor of engagement in coherent quiz activities ($b=-9.100$, $\beta=-0.283$, $p<.001$, 95% CI: $[-13.308, -4.890]$, $R^2=0.080$, adjusted $R^2=0.076$, $f^2=0.087$). This result implies that higher levels of pretest calibration discrepancy—an overestimation of their pretest performance—

were associated with reduced engagement in coherent quiz activities. However, pretest calibration discrepancy did not significantly predict engagement in coherent reading activity ($b = -0.074$, $\beta = -0.010$, $p = .891$, 95% $CI = [-1.132, 0.984]$, $R^2 = 0.009$, adjusted $R^2 = -0.005$, $f^2 = 0.001$).

RQ2. How are students' pretest retrospective judgments and actual performances separately related to their metacognitive strategy use?

Building upon the findings of RQ1, where students' pretest calibration discrepancy was a significant negative predictor of engagement in coherent quiz activities, RQ2 examines whether it is the students' perceived performance (i.e., pretest retrospective judgments) or their actual pretest scores that better predict the engagement in metacognitive strategies. To address RQ2, we conducted two linear regression analyses similar to those in RQ1. One model uses a coherent quiz as the dependent variable, while the other uses coherent reading. In both models, the independent variables are pretest retrospective judgment and actual pretest grade (Table 3). Our findings revealed that pretest retrospective judgments were significant negative predictors of engagement in coherent quiz activities ($b = -9.841$, $\beta = -0.306$, $p < .001$, 95% $CI: [-14.663, -5.019]$, $R^2 = 0.082$, adjusted $R^2 = 0.073$, $f^2 = 0.078$). This suggests that students who believed they performed well on the pretest were less likely to engage in quiz-based metacognitive strategies during the learning session. In contrast, actual pretest grades did not significantly predict engagement in coherent quiz activities ($b = -1.654$, $\beta = -0.048$, $p = .533$, 95% $CI: [-6.875, 3.567]$, $R^2 = 0.073$, adjusted $R^2 = 0.073$, $f^2 = 0.002$). Similarly, in the model examining coherent reading engagement, neither pretest retrospective judgments nor actual pretest grades were significant predictors. This aligns with the findings from RQ1, where pretest calibration discrepancy did not significantly predict engagement in coherent reading activities.

RQ3. How do SRL patterns, number of topic changes, and number of activity measures predict changes in students' metacognitive judgments?

In RQ3, we examined how learning patterns that are potentially associated with SRL-relevant strategies, the number of topic transitions, and the count of learning activities predict changes in students' metacognitive judgments. Understanding these

Table 3 Regression analysis for predictors of engagement in coherent quiz and reading (RQ2)

	Dependent variable	Independent variable	Coefficient	<i>p</i> -value	95% Confidence Interval
Coherent quiz and reading indicate students' engagement in those activities in minutes. Pretest retrospective judgment refers to students' estimated pretest grade	Coherent Quiz	Pretest retrospective judgment	-9.841	<0.001	[-14.663, -5.019]
		Actual pretest grade	-1.654	0.533	[-6.875, 3.567]
	Coherent Reading	Pretest retrospective judgment	-0.007	0.905	[-1.287, 1.140]
		Actual pretest grade	0.001	0.999	[-1.313, 1.314]

relationships provides valuable insights into which factors significantly contribute to adjustments in students' self-assessments after the learning session, should any be identified as significant predictors. To explore this, we conducted a linear regression analysis with the change in metacognitive judgments as the dependent variable. The independent variables included: (i) the counts of six learning patterns (as shown in the left-most column of Table 1), (ii) the total number of topic transitions, and (iii) the aggregate count of learning activities. Our analysis revealed that the frequency of the Quiz → Quiz sequence, potentially indicative of rehearsing and memorizing and repeating SRL-relevant strategy, was the only significant negative predictor of change in metacognitive judgments among the variables considered ($b = -1.792$, $p = .036$, 95% CI: $[-3.464, -0.120]$). This result suggests that an increase in students' engagement with the Quiz → Quiz pattern is associated with a decrease in the change in metacognitive judgment. In other words, students who frequently engaged in consecutive quizzes tended to experience less positive or even negative adjustments in their self-assessed performance after the learning session (Table 4).

Discussion

RQ1. How do students' pretest calibration discrepancies relate to their use of metacognitive strategies?

We observed relatively accurate calibration means, with calibration discrepancy measures of 0.796% points at pretest and 3.89% points at posttest. While our study design does not allow us to directly examine moderators such as the increased accuracy of delayed versus immediate judgments, prior work suggests that such factors may contribute to the observed outcomes (Nelson & Dunlosky, 1991). One possible explana-

Table 4 Regression analysis for factors influencing change in metacognitive judgment (RQ3)

Dependent variable	Independent variable	Coefficient	<i>p</i> -value	95% Confidence Interval
Change in meta-cognitive judgement	Read → Quiz	-0.258	0.831	$[-2.631, 2.116]$
	Quiz → Read	0.342	0.793	$[-2.227, 2.911]$
	Quiz → Quiz	-1.792	0.036*	$[-3.464, -0.120]$
	Quiz → Examples	-0.871	0.532	$[-3.616, 1.874]$
	Read → Examples	-0.750	0.538	$[-3.148, 1.647]$
	Quiz → Summary	-0.089	0.946	$[-2.670, 2.492]$
	Topic transitions	-0.678	0.148	$[-1.598, 0.243]$
	Learning activities	0.686	0.131	$[-0.206, 1.579]$

Topic transitions indicate the total number of topic transitions, and learning activities imply the total count of learning activities

tion is that taking the quizzes during the learning session, which allowed students to see which questions they answered incorrectly, may have helped them calibrate more accurately for the posttest. In our study, students also completed a pretest before the learning session, although they did not receive feedback on their answers. While this differs from typical pretesting implementations that provide feedback, future studies could investigate these moderators to better understand their impact on calibration accuracy.

Our findings revealed a significant negative relationship between pretest calibration discrepancy and subsequent engagement in coherent quiz ($b = -9.100$, $p < .001$, 95% CI : $[-13.308, -4.890]$) and support the hypothesis. This result suggests that students who tend to overestimate their knowledge (i.e., higher positive calibration discrepancy) are less likely to engage in coherent quiz activity, which involves proactive engagement with learning materials before attempting a quiz. This finding aligns with prior studies (Hadwin & Webster, 2013; Hattie, 2013) and adds empirical evidence that this relationship also holds in the computer-based learning environments examined in our study.

One possible explanation for this decreased engagement in coherent quiz is misplaced confidence: overconfident students may perceive themselves as adequately prepared and, therefore, feel little need to put in additional time to study before assessing their knowledge. This overconfidence may lead these students to skip essential preparatory steps, such as reviewing relevant materials, which are critical for ensuring a solid understanding of the content. As a result, these students may miss opportunities to enhance their understanding and adjust their learning strategies effectively during the learning process due to their misplaced overconfidence.

While much of the literature emphasize the risks of overconfidence (e.g., the Dunning–Kruger effect; Kruger & Dunning, 1999), which can lead students to adopt ineffective learning strategies and reduce time and effort devoted to studying (Aghababayan et al., 2017; Dunlosky & Rawson, 2012), underconfidence—reflected as negative calibration discrepancy in this study—may also impact students' effective learning. Specifically, our findings showed that students with lower calibration discrepancy values, including those who were accurately calibrated or underconfident, tended to spend more time in coherent quiz activity. This pattern may reflect a compensatory strategy in which students who doubt their knowledge invest additional effort in preparation. However, persistent underconfidence could also lead to inefficiencies, such as overstudying material that has already been mastered or experiencing unnecessary anxiety.

On the other hand, the lack of a significant relationship between pretest calibration discrepancy and coherent reading may be attributed to the reactive nature of the coherent reading activity. Coherent reading occurs after students receive immediate feedback from a quiz, which may prompt even students with misplaced confidence (e.g., overconfidence) to recognize the discrepancies between their perceived and actual performance. The quiz results serve as external indicators of their understanding, potentially motivating students to engage in remedial actions, such as reviewing missed material, regardless of their initial confidence levels. We further discuss the practical implications of our results in more detail in the *Practical Implications* subsection.

RQ2. How are students' pretest retrospective judgments and actual performances separately related to their metacognitive strategy use?

RQ2 found that students' pretest retrospective judgments were significant negative predictors of their engagement in coherent quiz activities ($b = -9.841$, $p < .001$, 95% CI: $[-14.663, -5.019]$), which differs from our hypothesis that expected students' actual performance would be a positive predictor. Our finding suggests that as students' estimated pretest scores increase, their engagement in coherent quiz decreases. In other words, students who believe they performed well on the pretest are less likely to invest time in preparatory activities before attempting quizzes during the learning session. Interestingly, students' actual pretest grades did not significantly predict their engagement in coherent quiz activities, indicating that students' decisions on how much time to spend preparing before assessments might be more strongly guided by students' subjective judgment of their knowledge and readiness rather than by their objectively measured performance. As a result, students who perceive themselves as well-prepared or knowledgeable, regardless of their actual performance, may be less inclined to put in additional effort in preparation before taking quizzes. This could be detrimental if students do not receive timely feedback to correct any potential early calibration discrepancies, especially in computer-based learning environments where there is limited instructor feedback. Without timely interventions, students may continue to operate under calibration discrepancies in self-assessments, which can hinder employing effective learning strategies, leading to suboptimal learning outcomes.

These results from RQ2 complement those of RQ1, emphasizing the critical role of students' metacognitive judgments early in the learning process in shaping their engagement with metacognitive strategies. The findings from RQ1 and RQ2 together underscore the need for timely correction of students' early calibration discrepancies—whether overconfidence or underconfidence—in computer-based learning environments. Addressing these early calibration discrepancies could help students engage more effectively in metacognitive strategies, ultimately leading to improved learning outcomes. Furthermore, these results call for further exploration into how students form these metacognitive judgments and the factors that influence their accuracy. Understanding these underlying factors could be crucial for designing AI tools that support students in making more accurate judgments of their knowledge, thereby fostering effective use of learning strategies and enhanced learning outcomes.

RQ3. How do SRL patterns, number of topic changes, and number of activity measures predict changes in students' metacognitive judgments?

Our analysis for RQ3 found that the frequency of the Quiz → Quiz learning pattern—likely reflecting rehearsing, memorizing, and repeating SRL strategies—was the only significant negative predictor of changes in metacognitive judgment ($b = -1.792$, $p = .036$, 95% CI: $[-3.464, -0.120]$). This finding suggests that increased engagement in the Quiz → Quiz pattern is associated with a decrease in the change in metacognitive judgment. In other words, frequent engagement in Quiz → Quiz decreased students' perceptions of how much they learned. This finding aligns with the underconfidence-with-practice effect (Koriat et al., 2002), which showed that

learners' judgments of learning tend to shift from slight overconfidence to underconfidence over repeated study–test cycles. In our context, repeated quiz-taking may have reinforced this underconfidence, as students became increasingly aware of the gaps in their knowledge.

Even though students felt less confident after the self-paced learning (as seen in their lower posttest predictions), engaging more in the Quiz → Quiz pattern seems to intensify this decline in perceived performance. In our study, students did not receive any feedback on their performance during the learning session unless they proactively took a quiz, which served as the only immediate form of feedback available within the learning system. Therefore, the reduced confidence observed in both cases could be attributed to the immediate feedback provided by the quizzes. It is likely that students who actively engaged in the Quiz → Quiz learning pattern received frequent feedback on questions that they got incorrect, which may have highlighted gaps in their understanding, leading to adjustments in their confidence levels more conservatively.

Students with strong SRL skills are more likely to engage in strategic self-evaluation, such as taking quizzes to assess their understanding of the material when they feel it is necessary, even without external assistance or interventions in computer-based learning environments. In contrast, students with weaker SRL skills are less likely to adopt these learning strategies, resulting in missed opportunities to recalibrate their inaccurate judgment of their performances. Such missed recalibration opportunities can lead to ineffective learning outcomes, as students may continue with a misplaced judgment of their performance. These RQ3 findings highlight the necessity for not only timely interventions to correct students' early calibration discrepancy, as identified in RQ1 and RQ2, but also for proactive support that guides students on how to recalibrate their misplaced judgments of their performance. This is because when students receive support with accurate performance information alone, they may lack the knowledge or skills to effectively utilize this information to modify their learning approaches. For instance, if overconfident students receive feedback indicating that their actual performance score is lower than anticipated, this information can help them identify the gap between their perceived and actual understanding. However, without guidance on how to address this discrepancy, students might not know how to adjust their study strategies or improve their comprehension of the material. Therefore, interventions should extend beyond merely presenting performance data to include structured guidance on interpreting feedback and implementing remedial strategies. AI-driven metacognitive calibration tools can play a pivotal role by offering personalized recommendations and strategy suggestions, enabling students to correct their pretest calibration discrepancies and engage in more effective self-regulated learning.

Practical Implications

Our findings offer valuable practical implications for enhancing existing calibration support tools in computer-based learning environments. First, the results indicating that pretest calibration discrepancy and metacognitive judgments are significant predictors of students' subsequent use of metacognitive strategies highlight the poten-

tial of using these metacognitive measures to enhance interventions by predicting students' metacognitive strategy use. The existing calibration supporting tools have potential to enhance their effectiveness by expanding their scope beyond merely offering feedback on the discrepancy between students' estimated and actual performance. For instance, overconfidence in performance at the onset of learning may lead to students' reduced engagement in metacognitive strategies. However, interventions could help students become aware of their calibration discrepancies by highlighting performance discrepancies. Furthermore, these interventions could assist students in leveraging this awareness, enabling them to refine their approaches to learning tasks and assessments through targeted suggestions that promote engagement with specific metacognitive strategies. Addressing students' overconfidence through a complementary approach—providing both performance discrepancy feedback and corresponding suggestions on the use of metacognitive strategies—can guide students in refining their learning strategies to correct calibration discrepancies.

Second, we emphasize the importance of early calibration interventions in online learning environments. Our results from RQ2 indicate that students' significantly predict their subsequent engagement with metacognitive skills. Thus, implementing calibration interventions after an assessment might miss key opportunities to influence students' learning strategies and behaviors at an earlier, potentially more impactful stage, prior to starting assessments. For instance, if calibration interventions provide feedback after the assessment, overconfident students before starting the assessment have missed the chance to employ effective learning strategies that could have improved their learning outcomes. However, early interventions could provide students with timely insights into their performance discrepancies and use of metacognitive strategies, helping them adjust their strategies before engaging deeply in learning tasks or assessments. This proactive approach could lead to more effective learning experiences by aligning students' self-perceptions with their actual performance early, thereby enhancing their overall engagement and effectiveness in learning activities.

Transitioning to early interventions could be particularly beneficial in online learning settings, where students often have to regulate their learning independently without constant guidance from instructors. Instead of solely focusing on post-assessment calibration intervention on discrepancy feedback, there are opportunities for integrating metacognitive support at the beginning of the learning process. Such an approach would not only correct students' initial misjudgments but also empower them to make more informed and effective decisions about their learning strategies throughout their learning in computer-based learning environments. Together, our findings not only advocate for a shift in the approach and timing of tools supporting metacognitive calibration support in computer-based learning environments but also underscore the fundamental importance of introducing such support to students more broadly. We argue that, given the fact that very few online learning systems currently offer explicit support for calibration, establishing a foundation for this kind of assistance is critical and necessary step, even if the tools seem less novel.

Limitations

While our study contributes to understanding the relationship between calibration discrepancy and learning behaviors, further research is needed to explore how these insights can be practically implemented to enhance learning outcomes in digital learning contexts. The scope of this study is limited, as the implications derived from it may not be fully applicable to online learning environments that lack early assessments before student learning sessions begin. The absence of initial assessments in some online learning environments necessitates further exploration into methods for effectively measuring or predicting students' early evaluations. Such research is vital for augmenting the impact of early interventions in these computer-based educational settings. Additionally, we acknowledge the limitation of using a single confidence judgment per test (pre/post) rather than item-level confidence ratings. Relying on a single, overall confidence rating may reduce the sensitivity of our calibration measure and prevents analysis of resolution (i.e., relative monitoring accuracy). Future work could benefit from collecting item-level confidence judgments to enable more fine-grained analyses of metacognitive monitoring.

Relatedly, although students predicted their overall test score (0–100%) rather than rating each item, such estimates may still be influenced by heuristics or anchors. Higham et al. (2015) note that numerical confidence judgments can be biased by psychological anchors. Future work could explore alternative measures of metacognitive calibration that reduce reliance on explicit scaling. Furthermore, we acknowledge that our study has limitations in providing causal evidence regarding whether students' calibration discrepancies or retrospective judgments influence their subsequent use of metacognitive strategies or if a common cause exists. Therefore, it is necessary to further examine the causal relationship between calibration discrepancy and strategy use through experimental designs and longitudinal studies. Such research could focus on developing targeted interventions to address students' calibration discrepancy and observe subsequent changes in metacognitive strategy use.

Conclusions

It is well-established that metacognitive skills are valuable for learning in computer-based learning environments (Efklides, 2006a; Pintrich et al., 2000; Zimmerman & Moylan, 2009), yet the dynamics between various facets of metacognition still require further examination (Zhao & Ye, 2020). This study contributes to understanding the complex relationships between initial metacognitive calibration, metacognitive strategy use, judgments, and academic performance using a theory-based method. In particular, our study uncovered the impact of pretest calibration discrepancies and judgments on students' use of metacognitive strategies. Our findings highlight the critical role of timely intervention in metacognitive support in computer-based learning environments. We advocate for a shift towards early, proactive calibration supporting tools, ensuring students' self-perceptions align with their actual performance early in their learning. Further, recognizing the significance of students' metacognitive judgments in their engagement with learning strategies, our study suggests

Table 5 Correlation between predictors in the regression model for RQ2.

	Pretest retrospective judgment	Actual pretest grade
Pretest retrospective judgment (1.309)	-	
Actual pretest grade (1.309)	-0.486	-

Pearson's r is reported for correlation. Values in parentheses are the variance inflation factors for each predictor

Table 6 Correlation between predictors in the regression model for RQ3.

	Read → Quiz count	Quiz → Read count	Quiz → Quiz count	Quiz → Ex-amples count	Read → Ex-amples count	Quiz → Summary count	Topic transitions count	Learning activities count
Read → Quiz count (2.534)	-							
Quiz → Read count (2.422)	0.653	-						
Quiz → Quiz count (1.924)	0.075	0.017	-					
Quiz → Examples count (2.125)	0.393	0.189	0.194	-				
Read → Examples count (2.300)	0.203	0.228	-0.020	0.483	-			
Quiz → Summary count (2.061)	0.277	-0.022	0.141	0.477	0.175	-		
Topic transitions count (5.944)	0.031	0.056	0.186	0.113	0.283	0.143	-	
Learning activities count (4.507)	0.292	0.259	0.430	0.463	0.519	0.449	0.683	-

Spearman's rho is reported for all correlations. Values in parentheses are the variance inflation factors for each predictor.

opportunities for calibration tools to offer more than mere identification of calibration inaccuracies to students. However, there remains a pressing need for more comprehensive studies aimed at unraveling these relationships in greater detail, particularly to gain a better understanding of how students' metacognitive judgments evolve throughout the learning process.

Appendix

Author Contributions HL and NB contributed to the conceptualization and methodology of this research. HL conducted the formal analysis, curated the data, and drafted the original manuscript. NB provided supervision and was involved in reviewing and editing the manuscript.

Funding The authors have no funding to report.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdelrahman, R. M. (2020). Metacognitive awareness and academic motivation and their impact on academic achievement of Ajman university students. *Heliyon*, 6(9), Article e04192. <https://doi.org/10.1016/j.heliyon.2020.e04192>
- Acosta-Gonzaga, E., & Ramirez-Arellano, A. (2021). The influence of motivation, emotions, cognition, and metacognition on students' learning performance: A comparative study in higher education in blended and traditional contexts. *SAGE Open*, 11(2), 215824402110275. <https://doi.org/10.1177/21582440211027561>
- Aghababayan, A., Lewkow, N., & Baker, R. (2017). Exploring the asymmetry of metacognition. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 115–119. <https://doi.org/10.1145/3027385.3027388>
- Akturk, A. O., & Sahin, I. (2011). Literature review on metacognition and its measurement. *Procedia - Social and Behavioral Sciences*, 15, 3731–3736. <https://doi.org/10.1016/j.sbspro.2011.04.364>
- Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, 24, 1–3. <https://doi.org/10.1016/j.learninstruc.2012.10.003>
- Aloqleh, A. M. A., & Teh, K. S. M. (2019). The effectiveness of metacognition on academic achievement among the Jordanian universities students. *International Journal of Academic Research in Business and Social Sciences*, 9(9), 460–478. <https://doi.org/10.6007/IJARBSS/v9-i9/6315>
- Avargil, S., Lavi, R., & Dori, Y. J. (2018). Students' metacognition and metacognitive strategies in science education. In Y. J. Dori, Z. R. Mevarech, & D. R. Baker (Eds.), *Cognition, Metacognition, and Culture in STEM Education: Learning, Teaching and Assessment* (pp. 33–64). Springer International Publishing. https://doi.org/10.1007/978-3-319-66659-4_3
- Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition And Learning*, 9(2), 217–228. <https://doi.org/10.1007/s11409-014-9123-1>
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M., Wortha, F., & Cerezo, R. (2022). Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2022.813632>

- Bogarín, A., Cerezo, R., & Romero, C. (2018). Discovering learning processes using inductive miner: A case study with learning management systems (LMSs). *Psicothema*, 30(3), 322–329. <https://doi.org/10.7334/psicothema2018.116>
- Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education*, 69(2), 133–151. <https://doi.org/10.1080/00220970109600653>
- Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, 73(4), 269–290. <https://doi.org/10.3200/JEXE.73.4.269-290>
- Bosch, N., Zhang, Y., Paquette, L., Baker, R., Ocumpaugh, J., & Biswas, G. (2021). Students' verbalized metacognition during computerized learning. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3411764.3445809>
- Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, 38(8), 941–956. <https://doi.org/10.1080/02602938.2013.769198>
- Bringula, R., Reguyal, J. J., Tan, D. D., & Ulfa, S. (2021). Mathematics self-concept and challenges of learners in an online learning environment during COVID-19 pandemic. *Smart Learning Environments*, 8(1), 22. <https://doi.org/10.1186/s40561-021-00168-5>
- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and Understanding* (pp. 65–116). Erlbaum.
- Cheng, E. C. K., & Chan, J. K. M. (2021). Metacognition and metacognitive learning. In E. C. K. Cheng & J. K. M. Chan (Eds.), *Developing Metacognitive Teaching Strategies Through Lesson Study* (pp. 11–24). Springer. https://doi.org/10.1007/978-981-16-5569-2_2
- Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, 14(1), 77–90. <https://doi.org/10.1016/j.lindif.2003.08.003>
- Cho, M. H., & Heron, M. L. (2015). Self-regulated learning: The role of motivation, emotion, and use of learning strategies in students' learning experiences in a self-paced online mathematics course. *Distance Education*, 36(1), 80–99. <https://doi.org/10.1080/01587919.2015.1019963>
- Chou, C. Y., Lai, K. R., Chao, P. Y., Lan, C. H., & Chen, T. H. (2015). Negotiation based adaptive learning sequences: Combining adaptivity and adaptability. *Computers & Education*, 88, 215–226. <https://doi.org/10.1016/j.compedu.2015.05.007>
- Chou, C. Y., & Zou, N. B. (2020). An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models. *International Journal of Educational Technology in Higher Education*, 17(1), Article 55. <https://doi.org/10.1186/s41239-020-00233-y>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Corrin, L., De Barba, P. G., & Bakharia, A. (2017). Using learning analytics to explore help-seeking learner profiles in MOOCs. *Proceedings of the 7th International Learning Analytics & Knowledge*, 424, 428. <https://doi.org/10.1145/3027385.3027448>
- Coutinho, S. A. (2007). The relationship between goals, metacognition, and academic success. *Educate*, 7, 39–47.
- Dumford, A. D., & Miller, A. L. (2018). Online learning in higher education: Exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education*, 30(3), 452–465. <https://doi.org/10.1007/s12528-018-9179-z>
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>
- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24, 58–61. <https://doi.org/10.1016/j.learninstruc.2012.05.002>
- Efklides, A. (2001). Metacognitive experiences in problem solving: Metacognition, motivation, and self-regulation. *Trends and prospects in motivation research* (pp. 297–323). Kluwer Academic.

- Efklides, A. (2006a). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3–14. <https://doi.org/10.1016/j.edurev.2005.11.001>
- Efklides, A. (2006b). Metacognitive experiences: The missing link in the self-regulated learning process. *Educational Psychology Review*, 18(3), 287–291. <https://doi.org/10.1007/s10648-006-9021-4>
- Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist*, 13(4), 277–287. <https://doi.org/10.1027/1016-9040.13.4.277>
- Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema*, 21(1), 76–82.
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, 46(1), 6–25. <https://doi.org/10.1080/00461520.2011.538645>
- Emory, B., & Luo, T. (2022). Metacognitive training and online community college students' learning calibration and performance. *Community College Journal of Research and Practice*, 46(4), 240–256. <https://doi.org/10.1080/10668926.2020.1841042>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition and Learning*, 12(1), 1–19. <https://doi.org/10.1007/s11409-016-9158-6>
- Goradia, T., & Bugarcic, A. (2017). A social cognitive view of self-regulated learning within online environment. *Advances in Integrative Medicine*, 4(1), 5–6. <https://doi.org/10.1016/j.aimed.2017.05.001>
- Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, 45(4), 203–209. <https://doi.org/10.1080/00461520.2010.515935>
- Hacker, D. J., & Bol, L. (2004). Metacognitive theory: Considering the social–cognitive influences. *Big theories revisited: Vol. 4: Research on Sociocultural influences on motivation and learning* (pp. 275–297). Information Age.
- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal Of Educational Psychology*, 92(1), 160–170. <https://doi.org/10.1037/0022-0663.92.1.160>
- Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning And Instruction*, 24, 37–47. <https://doi.org/10.1016/j.learninstruc.2012.10.001>
- Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. *Handbook of metacognition in education* (pp. 131–153). Routledge/Taylor & Francis Group.
- Hattie, J. (2013). Calibration and confidence: Where to next? *Learning and Instruction*, 24, 62–66. <https://doi.org/10.1016/j.learninstruc.2012.05.009>
- Helzer, E. G., & Dunning, D. (2012). Why and when peer prediction is superior to self-prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology*, 103(1), 38–53. <https://doi.org/10.1037/a0028124>
- Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2015). In J. Dunlosky, S. Uma, & K. Tauber (Eds.), *Internal mapping and its impact on measures of absolute and relative metacognitive accuracy* (Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199336746.013.15>
- Jalili, A., Hejazi, M., Foumani, E., G., & Morovati, Z. (2018). The relationship between meta-cognition and academic performance with mediation role of problem solving. *Quarterly Journal of Child Mental Health*, 5(1), 80–91.
- Kelemen, W. L., Winningham, R. G., & Weaver, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology*, 19(4–5), 689–717. <https://doi.org/10.1080/09541440701326170>
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education*, 104, 18–33. <https://doi.org/10.1016/j.compedu.2016.10.001>
- Klassen, R. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly*, 25(2), 88–102. <https://doi.org/10.2307/1511276>

- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131(2), 147–162. <https://doi.org/10.1037/0096-3445.131.2.147>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kubik, V., Koslowski, K., Schubert, T., & Aslan, A. (2022). Metacognitive judgments can potentiate new learning: The role of covert retrieval. *Metacognition And Learning*, 17(3), 1057–1077. <https://doi.org/10.1007/s11409-022-09307-w>
- Kuvalja, M., Verma, M., & Whitebread, D. (2014). Patterns of co-occurring non-verbal behaviour and self-directed speech; A comparison of three methodological approaches. *Metacognition And Learning*, 9(2), 87–111. <https://doi.org/10.1007/s11409-013-9106-7>
- Lee, H., & Bosch, N. (2024). Subtopic-specific heterogeneity in computer-based learning behaviors. *International Journal of STEM Education*, 11(1), 61. <https://doi.org/10.1186/s40594-024-00519-x>
- Lin, L. M., & Zabucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, 23(4), 345–391. <https://doi.org/10.1006/ceps.1998.0972>
- Liu Z., & Moon J. (2023). A framework for applying sequential data analytics to design personalized digital game-based learning for computing education. *Educational Technology & Society*, 26(2), 181–197. [https://doi.org/10.30191/ETS.202304_26\(2\).0013](https://doi.org/10.30191/ETS.202304_26(2).0013)
- Livingston, J. A. (2003). *Metacognition: An Overview* (ED474273). ERIC. <https://eric.ed.gov/?id=ED474273>
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15(1), 174–179. <https://doi.org/10.3758/PBR.15.1.174>
- Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 502–506. <https://doi.org/10.1037/a0021802>
- Mitsea, E., & Drigas, A. (2019). A journey into the metacognitive learning strategies. *International Journal of Online and Biomedical Engineering (iJOE)*. <https://doi.org/10.3991/ijoe.v15i14.11379>
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition And Learning*, 9(2), 75–85. <https://doi.org/10.1007/s11409-014-9114-2>
- Morphew, J. W. (2021). Changes in metacognitive monitoring accuracy in an introductory physics course. *Metacognition and Learning*, 16(1), 89–111. <https://doi.org/10.1007/s11409-020-09239-3>
- Narang, D., & Saini, S. (2013). Metacognition and academic performance of rural adolescents. *Studies on Home and Community Science*, 7(3), 167–175. <https://doi.org/10.1080/09737189.2013.11885409>
- Nederhand, M. L., Tabbers, H. K., & Rikers, R. M. J. P. (2019). Learning to calibrate: Providing standards to improve calibration accuracy for different performance levels. *Applied Cognitive Psychology*, 33(6), 1068–1079. <https://doi.org/10.1002/acp.3548>
- Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60053-5](https://doi.org/10.1016/S0079-7421(08)60053-5)
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science*, 2(4), 267–271. <https://doi.org/10.1111/j.1467-9280.1991.tb00147.x>
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal Of Experimental Education*, 74(1), 7–28.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition And Learning*, 1(2), 159–179. <https://doi.org/10.1007/s10409-006-9595-6>
- Osterhage, J. L., Usher, E. L., Douin, T. A., & Bailey, W. M. (2019). Opportunities for self-evaluation increase student calibration in an introductory biology course. *CBE—Life Sciences Education*, 18(2), ar16. <https://doi.org/10.1187/cbe.18-10-0202>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Paquette, L., Grant, T., Zhang, Y., Biswas, G., & Baker, R. S. (2021). Using epistemic networks to analyze self-regulated learning in an open-ended problem-solving environment. *Proceedings of the 2nd International Conference on Quantitative Ethnography*, 185–201. https://doi.org/10.1007/978-3-030-67788-6_13

- Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 54(4), 1643–1662. <https://doi.org/10.3758/s13428-021-01694-3>
- Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, 24(6), 1774–1784. <https://doi.org/10.3758/s13423-017-1242-7>
- Pesout, O., & Nietfeld, J. L. (2021). How creative am i? Examining judgments and predictors of creative performance. *Thinking Skills and Creativity*, 40, Article 100836. <https://doi.org/10.1016/j.tsc.2021.100836>
- Pieschl, S. (2009). Metacognitive calibration—An extended conceptualization and potential applications. *Metacognition And Learning*, 4(1), 3–31. <https://doi.org/10.1007/s11409-008-9030-4>
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. *Issues in the measurement of metacognition*. Buros Institute of Mental Measurements.
- Pradhan, S., & Das, P. (2021). Influence of metacognition on academic achievement and learning style of undergraduate students in Tezpur University. *Influence of Metacognition on Academic Achievement and Learning Style of Undergraduate Students in Tezpur University*, 10(1), 381–391.
- Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2022.913219>
- Saenz, G. D., Geraci, L., Miller, T. M., & Tirso, R. (2017). Metacognition in the classroom: The association between students' exam predictions and their desired grades. *Consciousness and Cognition*, 51, 125–139. <https://doi.org/10.1016/j.concog.2017.03.002>
- Saenz, G. D., Geraci, L., & Tirso, R. (2019). Improving metacognition: A comparison of interventions. *Applied Cognitive Psychology*, 33(5), 918–929. <https://doi.org/10.1002/acp.3556>
- Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers and Education: Artificial Intelligence*, 3, Article 100060. <https://doi.org/10.1016/j.caeai.2022.100060>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. <https://doi.org/10.1006/ceps.1994.1033>
- Segedy, J. R., Kinnebrew, J., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal Of Learning Analytics*, 2(1), 13–48. <https://doi.org/10.18608/jla.2015.21.3>
- Sobocinski, M., Malmberg, J., & Järvelä, S. (2017). Exploring temporal sequences of regulatory phases and associated interactions in low- and high-challenge collaborative learning sessions. *Metacognition and Learning*, 12(2), 275–294. <https://doi.org/10.1007/s11409-016-9167-5>
- Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics*, 2(1), 72–100. <https://doi.org/10.18608/jla.2015.21.5>
- Sorgenfrei, C., & Smolnik, S. (2016). The effectiveness of e-learning systems: A review of the empirical literature on learner control. *Decision Sciences Journal of Innovative Education*, 14(2), 154–184. <https://doi.org/10.1111/dsji.12095>
- Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE—Life Sciences Education*, 20(2), fe3. <https://doi.org/10.1187/cbe.20-12-0289>
- Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12(4), 437–475. <https://doi.org/10.1023/A:1009084430926>
- Sun, Q., Zhang, L. J., & Carter, S. (2021). Investigating students' metacognitive experiences: Insights from the English as a Foreign Language learners' writing metacognitive experiences questionnaire (EFLWMEQ). *Frontiers in Psychology*, 12, Article 744842.
- Taraban, R., Rynearson, K., & Kerr, M. S. (2000). Metacognition and freshman academic performance. *Journal of Developmental Education*, 24(1), 12.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal Of Educational Psychology*, 95(1), 66–73. <https://doi.org/10.1037/0022-0663.95.1.66>
- Tsai, Y., Lin, C., Hong, J., & Tai, K. (2018). The effects of metacognition on online learning interest and continuance to learn with MOOCs. *Computers & Education*, 121, 18–29. <https://doi.org/10.1016/j.compedu.2018.02.011>
- Urban, K., & Urban, M. (2019). Improving the accuracy of the self-evaluation during on-screen self-regulated learning through calibration feedback. *Proceedings of the 13th International Technology, Education and Development Conference*, 9002–9007. <https://doi.org/10.21125/inted.2019.2239>

- Van Der Stel, M., & Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: A longitudinal study from a developmental perspective. *European Journal of Psychology of Education*, 29(1), 117–137. <https://doi.org/10.1007/s10212-013-0190-5>
- Veenman, M. V. J., & van Cleef, D. (2019). Measuring metacognitive skills for mathematics: Students' self-reports versus on-line assessment methods. *ZDM Mathematics Education*, 51(4), 691–701. <https://doi.org/10.1007/s11858-018-1006-5>
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition And Learning*, 1(1), 3–14. <https://doi.org/10.1007/s11409-006-6893-0>
- Wagener, B. (2016). Metacognitive monitoring and academic performance in college. *College Teaching*, 64(2), 47–54. <https://doi.org/10.1080/87567555.2015.1116056>
- Ward, R. T., & Butler, D. L. (2019). An investigation of metacognitive awareness and academic performance in college freshmen. *Education*, 139(3), 120–126.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, 132(4), 408–428. <https://doi.org/10.3200/GENP.132.4.408-428>
- Winne, P. H., & Azevedo, R. (2014). Metacognition. In *The Cambridge handbook of the learning sciences*, 2nd ed. (pp. 63–87). Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.006>
- Winne, P. H., & Perry, N. E. (2000). Measuring Self-Regulated Learning. In *Handbook of Self-Regulation* (pp. 531–566). Elsevier. <https://doi.org/10.1016/B978-012109890-2/50045-7>
- Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education*, 140, Article 103595. <https://doi.org/10.1016/j.compedu.2019.103595>
- Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., Bosch, N., Munshi, A., & Biswas, G. (2020). The relationship between confusion and metacognitive strategies in Betty's Brain. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 276–284. <https://doi.org/10.1145/3375462.3375518>
- Zhang, Y., Paquette, L., Bosch, N., Ocumpaugh, J., Biswas, G., Hutt, S., & Baker, R. S. (2022). The evolution of metacognitive strategy use in an open-ended learning environment: Do prior domain knowledge and motivation play a role? *Contemporary Educational Psychology*, 69, Article 102064. <https://doi.org/10.1016/j.cedpsych.2022.102064>
- Zhao, L., & Ye, C. (2020). Time and performance in online learning: Applying the theoretical perspective of metacognition. *Decision Sciences Journal of Innovative Education*, 18(3), 435–455. <https://doi.org/10.1111/dsji.12216>
- Zhou, M. (2023). Students' metacognitive judgments in online search: A calibration study. *Education and Information Technologies*, 28(3), 2619–2638. <https://doi.org/10.1007/s10639-022-11217-y>
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal Of Educational Psychology*, 81(3), 329–339. <https://doi.org/10.1037/0022-0663.81.3.329>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. *Handbook of metacognition in education* (pp. 299–315). Routledge/Taylor & Francis Group.
- Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23(4), 614–628. <https://doi.org/10.2307/1163093>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

HaeJin Lee¹ · Nigel Bosch^{1,2}

✉ HaeJin Lee
haejin2@illinois.edu

¹ School of Information Sciences, University of Illinois Urbana-Champaign, 501 E. Daniel St., Champaign 61820, IL, USA

- ² Department of Educational Psychology, University of Illinois Urbana–Champaign, 1310 S 6th St, IL 61820 Champaign, USA