Fiegel, B., & Thomas, G. (2026). Gamified SQL Practice With Confidence Wagering: Links to Metacognition and Summative Performance. *ASEE Computers in Education*, *15*(3). https://doi.org/10.18260/B49A-99-68374

RESEARCH

Gamified SQL Practice With Confidence Wagering: Links to Metacognition and Summative Performance

Brian Fiegel, B.S., M.S.[1], Geb Thomas, B.S., M.S., Ph.D.[2]

[1] Industrial & Systems Engineering, University of Iowa, [2] Industrial And Systems Engineering, University of Iowa

Keywords: Engagement, Gamification, Deliberate practice, Metacognition

https://doi.org/10.18260/B49A-99-68374

ASEE Computers in Education

Vol. 15, Issue 3, 2026

Learning and retaining SQL can be challenging for undergraduate engineering students, particularly those early in their academic programs. The Deliberate Practice Engine is an adaptive e-learning platform designed to support SQL learning through individualized practice, immediate feedback, and a gamified wagering mechanism that prompts students to express confidence before submitting answers. This study examines whether wagering behavior provides insight into students' metacognitive monitoring and whether these behaviors predict performance on a summative SQL assessment.

Across two course offerings (N = 66), students generated 18,391 metacognitive responses. Wagering behavior was strongly bimodal, with most responses at 0 or 100 points. Accuracy increased with wager size: large wagers (100-point and All-In) were correct in 83.3% of attempts, compared to 59.5% for small wagers. Mixed-effects logistic regression confirmed that wager size was a robust predictor of correctness (OR = 2.84 for 100-point wagers; OR = 14.66 for All-In, both $p < .001$), even after accounting for question difficulty, topic level, format, response time, and student characteristics. These findings indicate that wagering functions as a meaningful behavioral proxy for metacognitive monitoring.

To evaluate whether metacognitive behavior predicted learning outcomes, we modeled exam performance separately for lower- and upper-division students. For lower-division students, a composite On-track metric, capturing metacognitive alignment and avoidance of overconfident errors, emerged as a significant predictor of SQL exam scores alongside prerequisite performance and quiz scores (adjusted $R^2 = 0.51$, $p < .05$). No practice-based predictors met selection criteria for upper-division students, whose exam scores were higher and less variable.

These results demonstrate that confidence wagering embedded within deliberate practice can reveal metacognitive processes and that such processes are especially informative for early-career engineering students. The findings highlight the value of integrating metacognitive scaffolds into adaptive practice systems and suggest design opportunities for supporting novice learners in technical domains.

1. Introduction

Undergraduate engineering students frequently struggle to learn and retain the structured query language (SQL), even after completing lectures, in-class exercises, and homework (Lahtinen et al., 2005; Robins et al., 2003). In our prior course offerings, students demonstrated rapid decay of SQL competence within weeks, a pattern consistent with research showing that declarative knowledge and syntax-heavy skills require repeated and varied practice to achieve fluency (Anderson, 2013; Cepeda et al., 2006; DeKeyser, 2020; Ericsson et al., 1993). To address this challenge, we developed the Deliberate Practice Engine, an adaptive e-learning platform that provides individualized SQL practice with immediate feedback and gamified elements. Earlier work demonstrated that the system substantially increases student engagement and supports improvement in SQL accuracy during practice sessions (Campbell, 2018; Fiegel et al., 2025; Yazvec, 2020).

This study extends our previous Deliberate Practice Engine work by examining whether the Engine's gamified confidence-wagering mechanism provides insight into students' metacognitive processes during practice, and whether those processes are linked to subsequent performance on a summative SQL assessment. Deliberate practice theory suggests that learners improve most efficiently when they repeatedly engage in progressively challenging tasks while monitoring their performance and adjusting their strategies accordingly (Anders Ericsson, 2008; Azevedo & Aleven, 2013; Hambrick et al., 2014; VanLehn, 2011). The Deliberate Practice Engine operationalizes this principle by adapting question difficulty to each learner's demonstrated proficiency and by offering iterative feedback on open-ended SQL queries. It also incorporates a point-based wagering feature that requires students to express their confidence before submitting an answer. Such confidence judgments are widely used as behavioral indicators of metacognitive monitoring, the ability to assess one's own knowledge state, and have been linked to improved self-regulated learning and higher academic achievement (Dunlosky & Metcalfe, 2008; Fleming, 2024; Koriat, 2006; Mahdavi, 2014; Nelson, 1990; Veenman et al., 2006).

Metacognition researchers distinguish between monitoring (evaluating what one knows) and control (regulating study behavior based on that evaluation) (Fischhoff et al., 1977; Keren, 1991; Nelson, 1990; Schraw & Dennison, 1994; Son & Schwartz, 2002). In learning environments that provide frequent practice opportunities, well-calibrated confidence judgments, high confidence when correct, low confidence when incorrect, are commonly interpreted as evidence of productive monitoring and control (Mayer & Alexander, 2017). Because the Deliberate Practice Engine captures explicit confidence signals on every initial attempt, it provides a unique opportunity to study how students calibrate their knowledge while learning a technical skill and whether those behaviors differ for novice and more advanced students.

This study investigates whether wager size functions as a meaningful proxy for metacognition and whether metacognitive wagering patterns help explain performance on a subsequent SQL exam. We analyze data from two consecutive offerings of a sophomore-level engineering course, examining 18,391 metacognitive responses generated by 66 students. Our analysis focuses on two related questions:

H1 (Metacognitive Monitoring)

Wager size will be positively associated with the likelihood of a correct response. Students who place larger wagers on items they answer correctly and smaller wagers on items they answer incorrectly demonstrate metacognitive alignment.

H2 (Performance Prediction)

Students who exhibit stronger metacognitive alignment during practice will perform better on the SQL portion of the course's summative assessment.

By linking students' moment-to-moment practice behavior to independent assessment outcomes, this study contributes to the growing body of work on gamification, metacognition, and deliberate practice in engineering education (Ishaq et al., 2024; Zeng et al., 2024). It also offers practical implications for designing instructional tools that support metacognitive development, particularly for early-career engineering students who may lack well-formed learning strategies.

2. Methods

Course Context and Participants

This study was approved by the University of Iowa Institutional Review Board (IRB #201908712). This was a retrospective study, and all data were de-identified prior to analysis.

This study analyzed data from two consecutive offerings of a sophomore-level Industrial Engineering course at a large Midwestern university. The course includes a two-week module introducing students to SQL through short lectures followed by extensive deliberate practice using the Deliberate Practice Engine. Sixty-six students participated and the course data was augmented with anonymized data for both the 2024 (n = 24) and 2025 (n = 42) offerings. First- and second-year students were classified as *lower-division* (n = 42), and third- and fourth-year students (plus a graduate students) as *upper-division* (n = 24). Participation with the system was part of normal coursework, and none of the analyses influenced students' grades.

Students were assigned weekly proficiency targets within the platform and completed a summative exam that included a common SQL component during week six of the course.

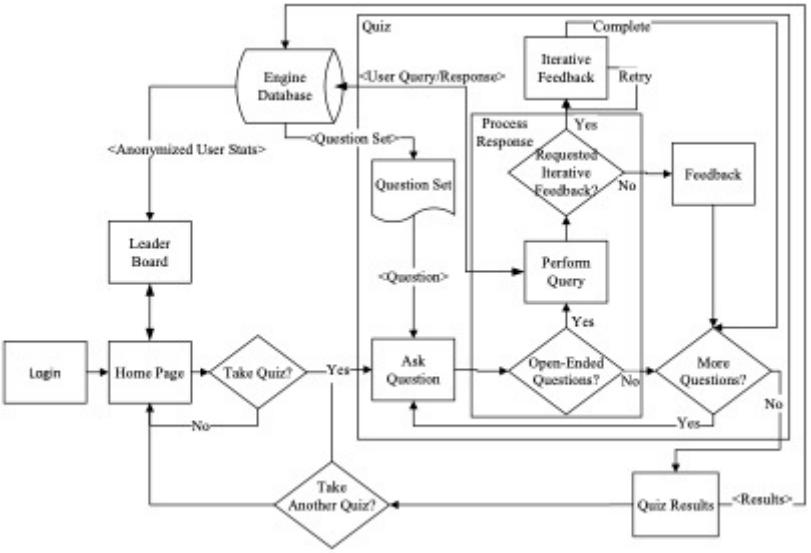


Figure 1. Flowchart depicts the student experience in the Deliberate Practice Engine.

The Deliberate Practice Engine

The Deliberate Practice Engine is an adaptive practice system that delivers SQL questions and provides students with immediate correctness feedback and opportunities for iterative refinement. Each quiz presents up to eight items drawn from a pool of 204 questions that span multiple SQL topics and three levels of conceptual progression (Basic, Intermediate, and Advanced). The system selects questions based on a student's current proficiency in each topic; *Figure 1* illustrates the overall flow from login to quiz initiation, to feedback and progression.

Questions vary in difficulty and format. Easy and intermediate questions are typically multiple-choice or true/false (Figure 2), while most hard questions require students to write complete SQL queries (Figure 3). For open-ended questions, students may request iterative feedback and iteratively run their query against the MySQL database. The system then returns the database output along with tailored hints designed to guide revision (Figure 4). Students may revise and resubmit open-ended answers until they choose to submit a final response. Because these iterative attempts are scaffolded and do not require students to express confidence, only the *first* attempt on each open-ended item is treated as a metacognitive response in our analyses.

A point-based wagering feature integrates a confidence judgment directly into practice. Before submitting each initial response, students may wager 0, 25, 50, 75, or 100 points; open-ended questions additionally allow an "All-In" wager that risks the student's entire point balance. Correct answers earn the wagered points along with a 10-point bonus, and incorrect answers lose the wagered points. Point totals reset to 0 if a loss exceeds the current balance. This wagering mechanism provides a behavioral proxy for metacognitive monitoring, because students indicate how confident they are in their answer before receiving feedback.



Figure 2. Sample Multiple Choice Question

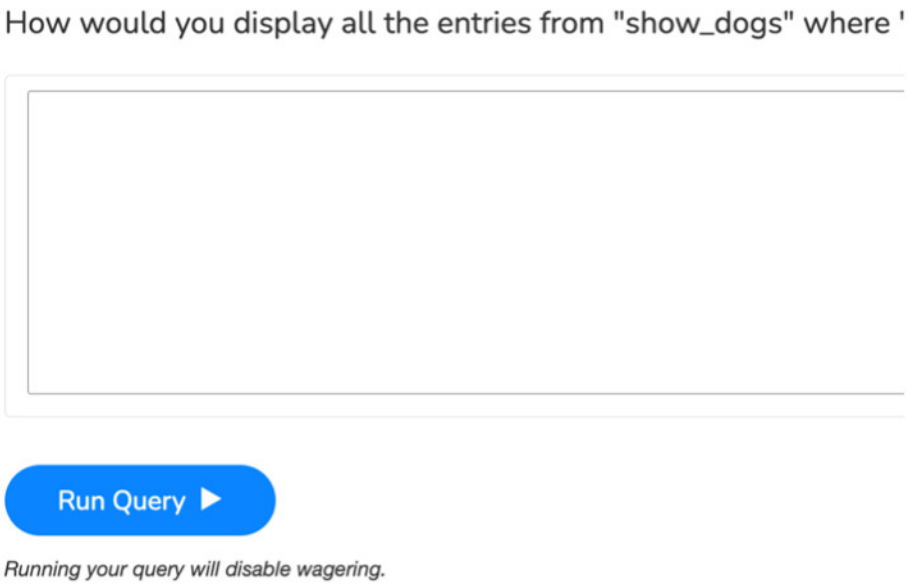


Figure 3. Sample Open-Ended Question

Data Sources and Measures

The Deliberate Practice Engine logs detailed information about each question attempt, including the student's wager, whether the response was correct, the response time, the question's difficulty level, the topic area, and whether the item was multiple-choice or open-ended. To capture only those attempts in which confidence was expressed, we analyzed all wagered responses as well as the first "No-Wager" attempt that initiates an iterative feedback cycle. Subsequent guided attempts within the iterative loop were excluded. Across both cohorts, the dataset included 18,391 such metacognitive responses.

From these logs we constructed several derived variables. Wager size was coded as *large* (100-point or All-In wagers) or *small* (0–75-point wagers and No-Wager entries). A response was coded as *metacognitively aligned* when students wagered large and were correct or wagered small and were incorrect.

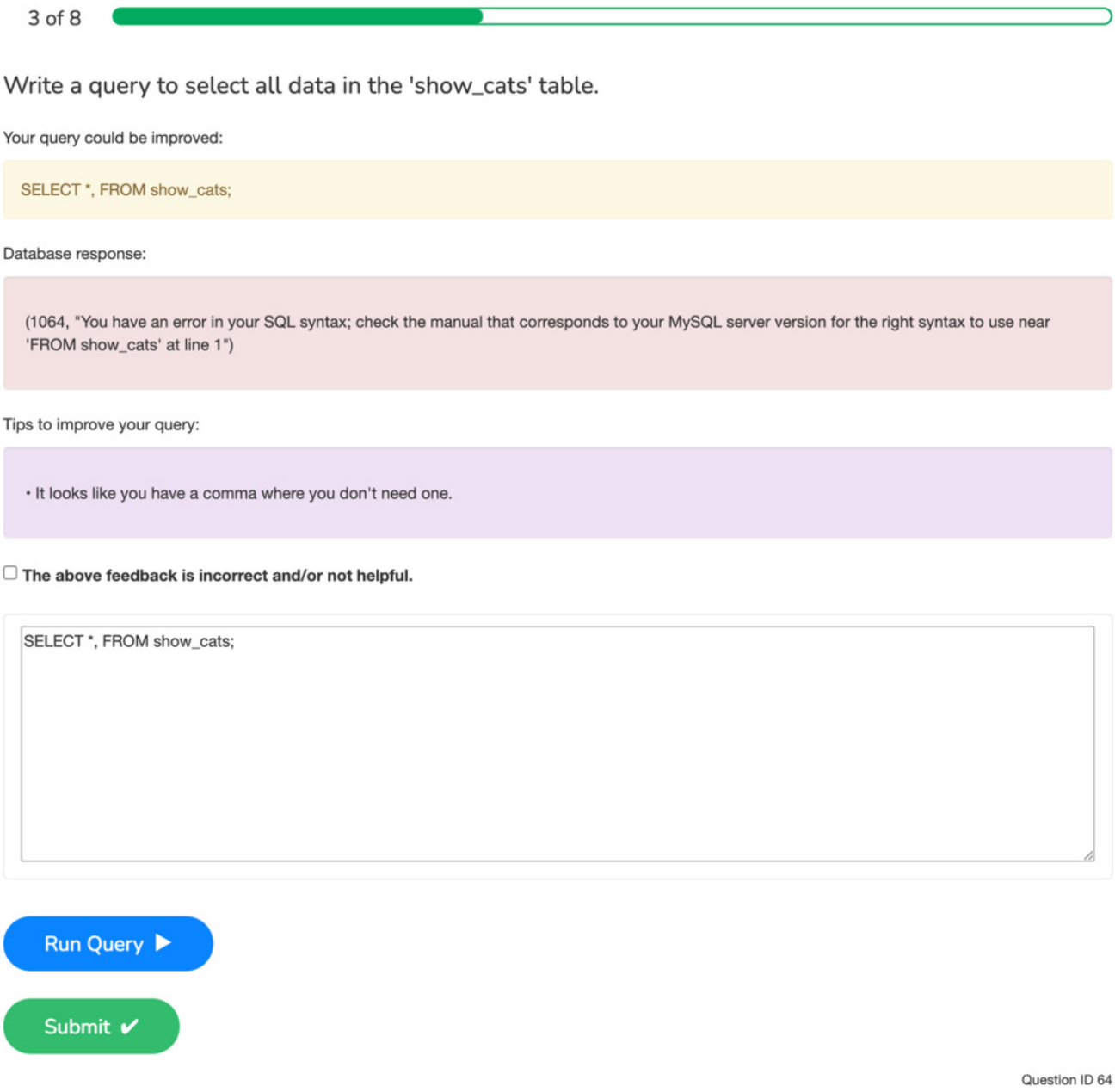


Figure 4. Sample feedback for an incorrect response to an open-ended question. The student's incorrect response is in the yellow section. The database response is in the pink section. The curated feedback text is in the purple section.

We also computed an *On-track* rate, defined as the proportion of aligned responses plus correct responses with small wagers—effectively capturing the extent to which students avoided overconfident errors.

Additional performance and demographic data were obtained from the course management system and university records. Course-level data included summative exam scores, two SQL quiz scores (combined into a standardized index), and homework scores (one of which was excluded due to near-zero variance). University record variables included cumulative GPA, earned engineering credit hours, the grade in the prerequisite programming course, and demographic indicators used in exploratory modeling. All institutional data were de-identified in accordance with IRB protocols.

Table 1. Student response statistics

Statistic	Value
Students (n)	66
Metacognitive responses	18391
Metacognitive responses per student, median (IQR) / mean	272.5(54.25) / 279
Metacognitive response pooled accuracy	0.711
Student level accuracy, mean (SD)	All 0.721 (0.086) Upper 0.763 (0.072) Lower 0.698 (0.085) Female 0.711 (0.087) Male 0.727 (0.086)
Responses by sex, count/mean (SD)	Female 6244/284 (64.05) Male 12147/276 (60.92)

Analytic Approach

Metacognitive Wagering and Response Accuracy (H1)

To test whether wager size predicted correctness after accounting for relevant contextual factors, we fit a mixed-effects logistic regression model with random intercepts for individual students. Fixed effects included wager category, question difficulty, question format, topic level, response time (log-transformed), division status (upper vs. lower), and correctness on the previous question within the same topic. This model assessed whether confidence judgments expressed through wagering served as reliable indicators of metacognitive monitoring.

Wagering Behavior and Summative Exam Performance (H2)

Because upper-division students scored substantially higher on the SQL exam, all analyses relating practice behavior to exam performance were conducted separately for the two cohorts. We first tested a targeted predictive model that included prerequisite course grade, the standardized quiz index, and alignment rate. To identify the most prominent predictors of exam performance, we applied LASSO stability selection across the full set of behavioral, course, and university-record variables, using 2,000 subsamples per cohort. Variables that met the cohort-specific selection frequency threshold were retained for the final regression model, which estimated the unique contribution of each predictor to exam performance while minimizing overfitting.

3. Results

Engagement and Performance

Sixty-six students generated 18,391 metacognitive responses (wagered responses plus first attempts in the iterative feedback loop). Students completed a median of 272.5 responses (IQR = 54.25; mean = 279). Pooled response accuracy was 71.1%; mean student-level accuracy was 72.1% (SD = 8.6%), with upper-division students outperforming lower-division students (76.3% vs. 69.8%; Table 1).

Table 2. Student SQL assessment statistics

Group	Students (n)	Mean exam score (SD)
All Studied Users	66	0.83
Upper-division	24	0.904 (.116)
Lower-division	42	0.781 (.164)
Female	22	0.850 (.130)
Male	44	0.814 (.172)

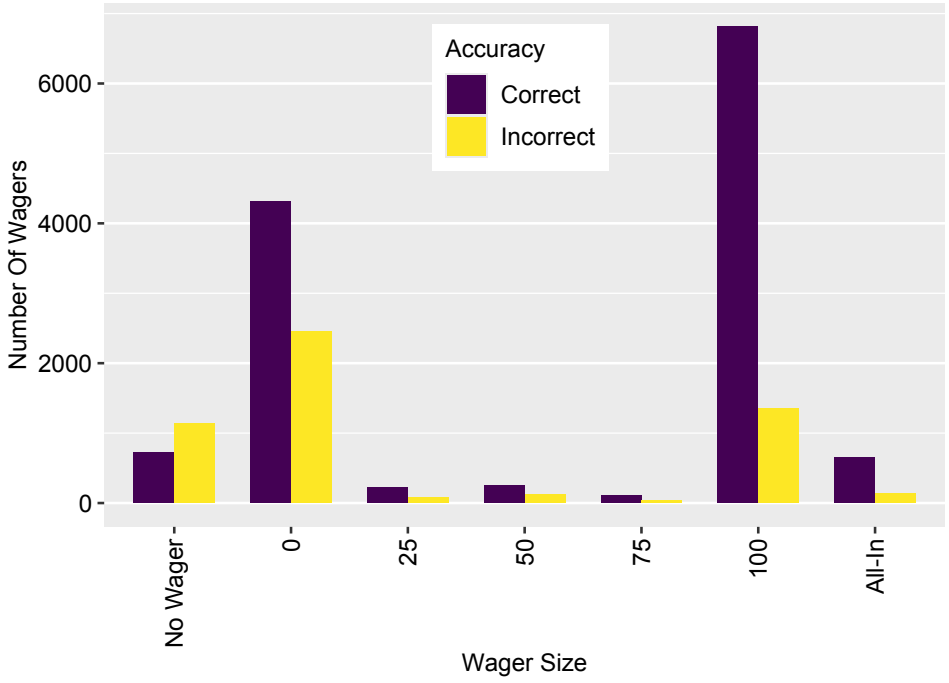


Figure 5. Wager value categorized by correct and incorrect answers.

On the SQL portion of the summative assessment, the overall mean score was 83%. Upper-division students scored 90.4% (SD = 11.6%) on average, compared to 78.1% (SD = 16.4%) for lower-division students (Welch's $t(61.0) = -3.55$, $p < 0.001$, Hedges' $g = 0.82$; Table 2).

Wagering Behavior and Metacognitive Alignment (H1)

Descriptive Patterns

Wagering behavior was strongly bimodal. Across all metacognitive responses, the distribution was: No Wager 10.1% (n = 1,851), 0 points 36.8% (n = 6,765), 25 points 1.7% (n = 304), 50 points 2.0% (n = 369), 75 points 0.8% (n = 138), 100 points 44.4% (n = 8,170), and All-In 4.3% (n = 794). Because intermediate wagers (25–75 points) were infrequent (4.5% in aggregate), they were collapsed into a <100 group for most analyses (Figure 5, Figure 6).

Accuracy increased with wager size. In the consolidated wager groups, percent correct was:



Figure 6. Wager breakdown (consolidated), by wager value, difficulty, and accuracy

Table 3. Accuracy by wager size

Wager Size	Attempts (n)	Correct (n)	Percent Correct	95% CI
Small	9427	5606	59.5%	58.5% to 60.5%
Large	8964	7469	83.3%	82.5% to 84.1%

- **No Wager:** 38.8% (95% CI [36.6, 41.1])
- **<100:** 64.5% (95% CI [63.4, 65.6])
- **100:** 83.4% (95% CI [82.6, 84.2])
- **All-In:** 82.7% (95% CI [80.0, 85.2])

When No Wager and <100 were grouped as Small and 100 and All-In as Large (Table 3), accuracy was 59.5% for Small wagers and 83.3% for Large wagers. Thus, large wagers were associated with substantially higher accuracy than small wagers.

Metacognitive alignment, defined as large wagers when correct and small wagers when incorrect, was also patterned by wager size and difficulty (Figure 7, Figure 8). Overall, 61.4% of responses were aligned and 38.6% unaligned. Large wagers were predominantly aligned across difficulty levels (≥ ~80%), whereas small wagers were more often unaligned on Easy and Intermediate questions but became more aligned on Hard items.

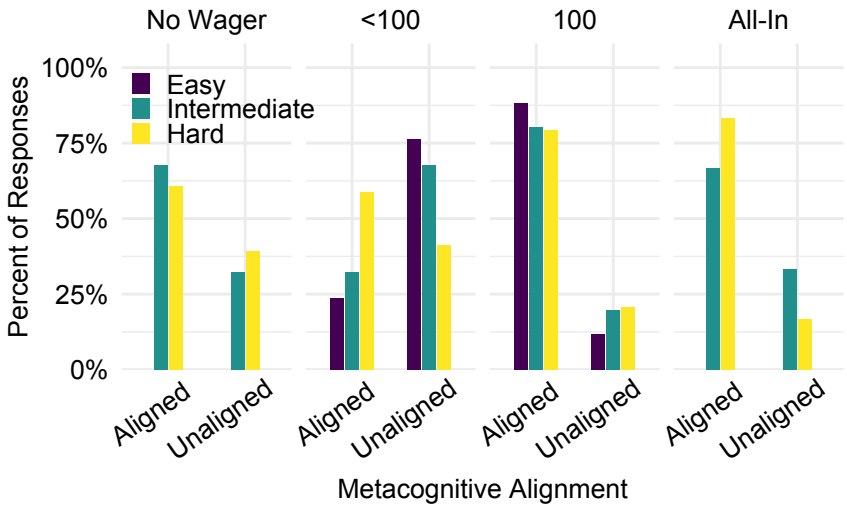


Figure 7. Metacognitive Alignment by wager and difficulty.

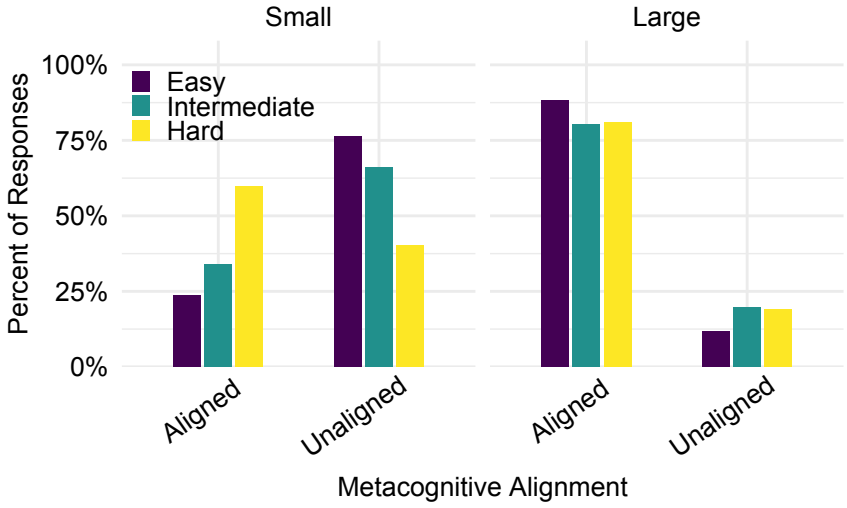


Figure 8. Metacognitive Alignment by wager size and difficulty.

Question-Level Modeling

To evaluate H1 while accounting for question and student characteristics, we fit a mixed-effects logistic regression predicting correctness from wager size and covariates (Table 4). The model was:

$$P(Y = 1) = 1/(1 + e^{(-f(X))})$$

where Y is the probability of a correct response and $f(X)$ is the linear predictor:

$$\begin{aligned} f(X) = {} & \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \\ & + \beta_6 X_6 + \beta_7 X_7 + \beta_8 (X_5 * X_6) + \beta_9 (X_5 * X_7) \\ & + \beta_{10} (X_6 * X_7) + u_i \end{aligned}$$

where:

X_1 is the correctness on the previous question in the same topic (0/1),
X_2 is the student division (lower vs. upper),
X_3 is the topic level (Basic, Intermediate, Advanced),
X_4 is the natural log of the response time,
X_5 is the Wager Size (No Wager, 0, 25, 50, 75, 100, or All-In),
X_6 is the Question Format (open-ended vs. multiple-choice),
X_7 is the Question Difficulty level (Easy, Intermediate, Hard),
$\beta_0 ... \beta_7$ are the regression coefficients,
$\beta_8, \beta_9, \beta_{10}$ are the interaction coefficients,
u_i is the random intercept for student i, $u_i \sim N(0, \sigma_u^2)$

Wager size remained a strong predictor of correctness after adjustment. Relative to 0-point wagers:

- 100-point wagers were associated with 2.84 times higher odds of correctness (95% CI [2.39, 3.37], p < .001).
- All-In wagers were associated with 14.66 times higher odds of correctness (95% CI [11.49, 18.70], p < .001).
- No Wager responses also showed modestly higher odds of correctness (OR = 1.33, 95% CI [1.11, 1.58], p = .002).

Intermediate wagers (25–75 points) did not show consistent significant main effects, reflecting their sparse and heterogeneous use.

Several covariates behaved as expected. The Temporal Dependency term indicates that a correct response on the previous question within the same topic increased the odds of correctness on the current response (OR = 1.43, 95% CI [1.32, 1.55], p < .001). Upper-division students had higher odds of correctness than lower-division students (OR = 1.42, 95% CI [1.11, 1.81], p = .005). Both Intermediate and Advanced topic levels showed higher odds of correctness than Basic, controlling for difficulty

Question difficulty and response time were also associated with performance. Hard questions substantially reduced the odds of correctness relative to Easy questions (OR = 0.44, 95% CI [0.36, 0.54], p < .001). Longer response times were associated with lower odds of correctness (OR = 0.82, 95% CI [0.78, 0.85], p < .001), consistent with longer deliberation reflecting uncertainty.

Format × difficulty interactions indicated that open-ended items became relatively more difficult than multiple-choice items at higher difficulty levels. Selected wager × difficulty interactions suggested that calibration varied somewhat by difficulty, but the overall pattern, higher wagers predicting higher correctness, was preserved.

Table 4. Mixed-Effects Logistic Regression Predicting Response Correctness with Full Controls (n = 17,053 observations from 66 students).

Fixed Effects	Odds Ratio	95% CI	p	Sig
Temporal Dependency				
Previous Response Correct	1.43	1.32 to 1.55	<0.001	***
Student Division (reference: Lower-division)				
Upper-division	1.42	1.11 to 1.81	0.005	**
Topic Level (reference: Basic)				
Intermediate	1.31	1.20 to 1.44	<0.001	***
Advanced	1.26	1.14 to 1.40	<0.001	***
Log(response time)	0.82	0.78 to 0.85	<0.001	***
Wager Size (reference: 0 points)				
No Wager	1.33	1.11 to 1.58	0.002	**
25 points	1.46	0.92 to 2.31	0.104	
50 points	0.98	0.67 to 1.42	0.903	
75 points	1.27	0.66 to 2.43	0.476	
100 points	2.84	2.39 to 3.37	<0.001	***
All-In	14.66	11.49 to 18.70	<0.001	***
Question Format (reference: Multiple-choice)				
Open-ended	1.36	1.02 to 1.82	0.036	*
Question Difficulty (reference Easy)				
Intermediate	0.87	0.74 to 1.03	0.107	
Hard	0.44	0.36 to 0.54	<0.001	***
Format x Difficulty Interaction				
Open-ended x Intermediate	0.40	0.30 to 0.54	<0.001	***
Open-ended x Hard	0.33	0.23 to 0.46	<0.001	***
Key Wager x Difficulty Interaction				
No Wager x Intermediate Difficulty	0.30	0.20 to 0.46	<0.001	***
All-In x Intermediate Difficulty	0.17	0.07 to 0.42	<0.001	***
25 points x Hard Difficulty	3.14	1.03 to 9.58	0.045	*
50 points x Hard Difficulty	4.78	1.83 to 12.50	0.001	**
100 point x Hard Difficulty	1.92	1.46 to 2.52	<0.001	***

Significance key: *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$, . $p \leq .1$

Model Fit: AIC=17330.2

Random Effects: Student-level variance $\sigma^2 = 0.21$ (SD = 0.45)

Note: Coefficients presented as Odds Ratio with 95% confidence interval.

Note: Model is rank-deficient as expected (No Wager and All-In only occur on open-ended questions; open-ended questions only occur at Intermediate/Hard difficulty). Four interaction coefficients representing impossible combinations were automatically omitted by lme4.

Wagering and SQL Summative Assessment Performance (H2)

Because upper-division students scored substantially higher on the SQL exam, we modeled exam performance separately for lower- and upper-division cohorts.

Lower-Division Students

For lower-division students (n = 42), we first tested Hypothesis 2 using a targeted linear regression

Table 5. Aligned Linear Model for SQL Summative Assessment, Lower-division Students.

Predictor	Estimate	SE	t	p	Sig.
β_1, Intro to Eng Computing Grade	0.056	0.017	3.38	0.002	**
β_2, Quiz Assessments	0.058	0.022	2.61	0.013	*
β_3, Alignment Rate	-0.319	0.137	-2.32	0.026	*

Significance key: *** p ≤ .001, ** p ≤ .01, * p ≤ .05, . p ≤ .1
Residual standard error: 0.124 on 38 degrees of freedom
Multiple R-squared: 0.473, Adjusted R-squared: 0.432
F-statistic: 11.37 on 3 and 38 DF, p-value: <0.001

$$y_i = \beta_o + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i$$

where:

y_i is the SQL exam score for student i,

β_o is the intercept,

β_1, β_2, β_3 are the regression coefficients,

x_1 is the z-scored prerequisite course grade (Intro to Eng Computing),

x_2 is the z-scored SQL quiz index (scaled mean on the two low-stakes practice quizzes),

x_3 is the Wilson-adjusted, centered, and scaled Alignment Rate (proportion of aligned responses among metacognitive responses),

ε_i is the error term for each student and is normally distributed with mean zero and variance σ^2, $\varepsilon_i \sim N\left(0, \sigma^2\right)$

In this initial model, alignment rate was negatively associated with exam performance (Table 5), contrary to the hypothesized direction.

In response to the negative association of alignment rate with exam performance and to better capture productive learning behavior, we defined an expanded metric: the On-track rate, comprising aligned responses (large-wager correct, small-wager incorrect) plus small-wager correct responses. Equivalently, On-track rate is 1 minus the rate of overconfident errors (large-wagers on incorrect responses).

We then applied LASSO stability selection over the full set of university-record, course, and Deliberate Practice Engine metrics. Features were ranked by selection frequency across 2,000 repetitions, with a selection frequency threshold of 0.65 and an upper bound of √n predictors. The resulting candidate set was used to build a final linear model for lower-division exam performance (Table 6).

The final model included:

- Intro to Eng Computing grade (β = 0.055, SE = 0.017, p = .002)
- SQL quiz index (β = 0.048, SE = 0.021, p = .029)
- On-track rate (β = 0.667, SE = 0.260, p = .015)

Table 6. Final Linear Model for SQL Summative Assessment, Lower-division Students.

Predictor	Estimate	SE	t	p	Sig.
Intro to Eng Computing Grade	0.055	0.017	3.26	0.002	***
Quiz Assessments	0.048	0.021	2.28	0.029	*
On Track Rate	0.667	0.260	2.57	0.015	*
Earned Engineering Credit Hours	0.013	0.006	2.35	0.025	*
Wager 75 Accuracy	0.287	0.231	1.24	0.221	

Significance key: *** p ≤ .001, ** p ≤ .01, * p ≤ .05, . p ≤ .1
Residual standard error: 0.1149 on 36 degrees of freedom
Multiple R-squared: 0.5695, Adjusted R-squared: 0.5098
F-statistic: 9.526 on 5 and 36 DF, p-value: <0.001

Table 7. Aligned Linear Model for SQL Summative Assessment, Upper-division Students.

Predictor	Estimate	SE	t	P	Sig.
Intro to Eng Computing Grade	0.046	0.024	1.869	0.0789	.
Quiz Assessments	0.016	0.021	0.791	0.4396	
Alignment Rate	0.024	0.187	0.129	0.8992	

Significance key: *** p ≤ .001, ** p ≤ .01, * p ≤ .05, . p ≤ .1
Residual standard error: 0.1045 on 17 degrees of freedom
Multiple R-squared: 0.3586, Adjusted R-squared: 0.2454
F-statistic: 3.168 on 3 and 17 DF, p-value: 0.05129

- Earned Engineering Credit Hours (β = 0.013, SE = 0.006, p = .025)
- Wager-75 Accuracy (not statistically significant)

This model explained a substantial portion of variance (adjusted $R^2 \approx 0.51$). The positive coefficient for On-track rate indicates that lower-division students who more frequently combined correct responses with appropriate wagering (including cautious correct responses) tended to achieve higher SQL exam scores.

Upper-Division Students

For upper-division students (n = 24), the same procedure was applied. The targeted alignment model (Table 7) did not yield significant effects of alignment rate, and LASSO stability selection did not identify any Deliberate Practice Engine metrics that met the selection threshold. Exam scores in this cohort were higher and less variable overall, which may have limited the ability of in-game behaviors to explain additional variance.

4. Discussion

This study examined whether a gamified deliberate practice platform with point-based wagering could serve as a behavioral probe of metacognition in undergraduate SQL learning and whether metacognitive wagering patterns

would predict independent assessment outcomes. Results from 18,391 metacognitive responses across two course offerings yielded two main findings.

First, wager size was robustly associated with correctness. Descriptive analyses revealed a strongly bimodal wagering distribution and substantial differences in accuracy across wager bands, with large wagers (100 and All-In) being accurate more than 80% of the time and small wagers under 60% of the time. Mixed-effects logistic regression showed that these wager effects persisted after controlling for question difficulty, format, topic progression, student division, prior correctness, and response time. These patterns support the interpretation of wagering as a behavioral indicator of metacognitive monitoring, in line with theoretical accounts in which confidence judgments track perceived knowledge states.

Second, metacognitive practice indicators were meaningfully related to exam performance for lower-division students, but not for upper-division students. For the lower-division cohort, exam performance was best explained by a combination of prior computing performance (Intro to Eng Computing grade), SQL quiz performance, earned engineering credit hours, and the derived On-track rate. The On-track metric, which rewards both accurate responding and avoidance of overconfident errors, behaved as hypothesized: higher On-track rates were associated with higher SQL exam scores. For upper-division students, whose exam performance was higher and less variable, Deliberate Practice Engine metrics did not emerge as robust predictors.

The choice to move from the initial wager alignment metric to the on-track metric requires qualification. The wager alignment metric treated wagers of 0 points as low-confidence responses. However, the default wager was zero, so students ignoring the wagering mechanism was indistinguishable from genuine low-confidence wagering. This conflation likely introduced measurement noise and may have contributed to the unexpected negative correlation between performance and the wager alignment metric. In contrast, the On-track metric more directly captured metacognitive calibration: it credited correct responses, tolerated incorrect responses paired with low confidence, and penalized incorrect high-confidence responses. This structure better aligned the metric with calibrated performance rather than raw confidence expression.

Upper-division students showed minimal effects for either measure, likely due to restricted variance in exam performance. Additionally, because the course is typically taken in the sophomore year, non-sophomore students constituted a self-selected subgroup whose prior academic experiences may have influenced wagering behavior independently of metacognitive skill.

Taken together, these findings indicate that gamified confidence judgments can be harnessed as informative metacognitive signals in authentic disciplinary practice. Wagering decisions in the Deliberate Practice Engine were not arbitrary or purely motivational; rather, they aligned closely with

correctness and shifted systematically with contextual and temporal information. Moreover, for students earlier in their programs, the way they combined accuracy and risk management during practice was associated with their performance on a separate, course-based assessment.

From a design perspective, the results suggest several implications for gamified deliberate practice systems:

1. **Coarse, high-stakes confidence scales may be sufficient.** The near absence of mid-range wagers and the strong calibration of large vs. small wagers suggest that a simplified wagering scheme can yield interpretable metacognitive data without requiring fine-grained confidence ratings.
2. **Composite metrics such as On-track are useful for instructional decision-making.** Because On-track integrates correctness and overconfidence, it could support dashboards that highlight students who are frequently overconfident (large wagers on incorrect answers) or persistently cautious despite high accuracy. Such analytics could enable targeted coaching around self-regulated learning.
3. **Metacognitive practice may be especially valuable for early-stage learners.** The fact that On-track predicted exam performance only for lower-division students suggests that metacognitive scaffolds embedded in deliberate practice may have their greatest impact when knowledge structures and learning habits are less established.

These contributions should be interpreted in light of several limitations. The study was conducted in a single course at one institution with a relatively small upper-division sample. The SQL module spanned a limited duration, and exam scores were only one indicator of learning. Wagering behavior may also be influenced by individual differences in risk preference, motivation, or familiarity with game mechanics. Although the models controlled for many observable factors, unmeasured variables may contribute to both wagering and performance.

Future research could examine how metacognitive wagering patterns evolve over longer courses or across multiple courses, how explicit instruction about wagering strategies affects calibration, and whether similar patterns appear in other domains (e.g., statistics, programming, or physics). It would also be informative to explore how wagering data can be integrated into real-time feedback, such as prompts triggered by overconfident errors.

5. Conclusion

The Deliberate Practice Engine's wagering feature provided a rich, behaviorally grounded view of students' metacognitive processes during SQL practice. Wager size was strongly and consistently related to correctness, both descriptively and in mixed-effects models, indicating that students used the wagering mechanism to express calibrated confidence rather than random guesses. Students also adjusted wagers based on recent performance, topic level, question format, and response fluency, demonstrating metacognitive regulation.

For lower-division students, a composite On-track metric that rewarded both accurate responding and avoidance of overconfident errors was positively associated with SQL exam performance, above and beyond prior coursework and quiz scores. For upper-division students, exam performance was not strongly predicted by Deliberate Practice Engine behaviors, likely reflecting higher baseline competence and reduced variability.

These results indicate that gamified deliberate practice with embedded wagering can support and reveal metacognitive processes in technical learning and that metacognitive practice features may be particularly beneficial for students early in their academic trajectories. Further, encouraging metacognitive alignment outside of gamified environments could benefit student performance, especially for lower division students.

Acknowledgements

We would like to acknowledge support from the University of Iowa Office of Teaching, Learning, and Technology (OTLT) Research & Analytics team. Their assistance with the IRB process and data preparation has been greatly appreciated.

We would also like to acknowledge the Hanson Center for Communication. Their assistance in keep progress focused is what allowed this article to be completed in a timely fashion.

ChatGPT 5.1 by OpenAI and Claude Sonnet 4.5 by Anthropic were used to 1) review and propose revisions to R code and 2) edit the prose for clarity and concision. All suggestions were checked by the authors, with code tested against the analysis outputs reported in this paper. The authors accept responsibility for the accuracy and integrity of all content. No AI generated content was accepted without human verification, and no AI system was used to generate data or figures.

Submitted: December 15, 2025 EDT. Revised: May 15, 2026 EDT. Accepted: June 25, 2026 EDT. Published: July 31, 2026 EDT.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.

References

Anders Ericsson, K. (2008). Deliberate Practice and Acquisition of Expert Performance: A General Overview. *Academic Emergency Medicine*, *15*(11), 988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Anderson, J. R. (2013). *Cognitive skills and their acquisition*. Psychology Press. https://doi.org/10.4324/9780203728178

Azevedo, R., & Aleven, V. (2013). *International handbook of metacognition and learning technologies* (Vol. 26). Springer. https://doi.org/10.1007/978-1-4419-5546-3

Campbell, S. L. (2018). *An evaluation of an automated, interactive learning method for a database query language*. University of Iowa.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354. https://doi.org/10.1037/0033-2909.132.3.354

DeKeyser, R. (2020). Skill acquisition theory. In *Theories in second language acquisition* (pp. 83–104). Routledge. https://doi.org/10.4324/9780429503986-5

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The Role of Eeliberate Practice in the Acquisition of Expert Performance. *Psychological Review*, *100*(3), 363. https://doi.org/10.1037/0033-295X.100.3.363

Fiegel, B. M., Thomas, G. W., & Yazvec, M. (2025). Enhancing SQL Learning Through Gamified Deliberate Practice: The Impact of Engagement and Feedback in Undergraduate Education. *2025 ASEE Annual Conference & Exposition*.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552.

Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology*, *75*(1), 241–268. https://doi.org/10.1146/annurev-psych-022423-032425

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*, *45*, 34–45. https://doi.org/10.1016/j.intell.2013.04.001

Ishaq, K., Alvi, A., ul Haq, M. I., Rosdi, F., Choudhry, A. N., Anjum, A., & Khan, F. A. (2024). Level up your coding: A systematic review of personalized, cognitive, and gamified learning in programming education. *PeerJ Computer Science*, *10*, e2310. https://doi.org/10.7717/peerj-cs.2310

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217–273. https://doi.org/10.1016/0001-6918(91)90036-Y

Koriat, A. (2006). *Metacognition and consciousness*. Institute of Information Processing and Decision Making, University of Haifa.

Lahtinen, E., Ala-Mutka, K., & Järvinen, H.-M. (2005). A study of the difficulties of novice programmers. *ACM SIGCSE Bulletin*, *37*(3), 14–18. https://doi.org/10.1145/1151954.1067453

Mahdavi, M. (2014). An Overview: Metacognition in Education. *International Journal of Multidisciplinary and Current Research*, *2*(6), 529–535.

Mayer, R. E., & Alexander, P. A. (2017). *Handbook of research on learning and instruction*. Routledge.

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

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, *13*(2), 137–172. https://doi.org/10.1076/csed.13.2.137.14200

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Son, L. K., & Schwartz, B. L. (2002). The relation between metacognitive monitoring and control. *Applied Metacognition*, 15–38. https://doi.org/10.1017/CBO9780511489976.003

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

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

Yazvec, M. (2020). *The effects of wagering and iterative feedback on engagement and performance in a MySQL learning context*. University Of Iowa.

Zeng, J., Sun, D., Looi, C. K., & Fan, A. C. W. (2024). Exploring the impact of gamification on students' academic performance: A comprehensive meta-analysis of studies from the year 2008 to 2023. *British Journal of Educational Technology*, *55*(6), 2478–2502. https://doi.org/10.1111/bjet.13471