2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC

Host Sponsor UNC Charlotte

Paper ID #50238

Enhancing Student Engagement and Confidence through the Instructor Mistake Meter

Andrew Michael Frisbie, Northeastern University
Lindsey Young, Northeastern University
Dr. Daniel Grindle, Northeastern University

Enhancing Student Engagement and Confidence through the Instructor Mistake Meter

Abstract

Active student engagement is crucial for effective learning, yet many students hesitate to participate in class due to fear of embarrassment or appearing unintelligent. The Instructor Mistake Meter (IMM) was developed to address this challenge by rewarding students with bonus points for identifying instructor errors during lectures, aiming to increase engagement and foster psychological safety. This study examined the IMM's effects on student participation, confidence, and classroom comfort across two semesters of an undergraduate biomechanics course. The experimental semester (Fall 2025, n=67) used the IMM, while the control semester (Spring 2025, n=62) received equivalent bonus points through an alternative assignment. Data collection included question-tracking during lectures, anonymous surveys assessing academic confidence and classroom comfort, and qualitative reflections from experimental students.

Results showed that experimental students asked significantly more questions than control students, with a 139% increase in daily question rates (p=0.001). Thematic analysis of qualitative responses revealed six primary themes, with student engagement (54% of responses) and motivation (46%) most frequently cited. Students reported increased attentiveness, comfort asking questions, and improved learning through error identification. However, quantitative surveys showed no significant differences in self-reported confidence or comfort between groups. While most student feedback was positive, some noted concerns about trivial errors becoming distracting. These findings suggest that the IMM effectively increases classroom participation and may enhance psychological safety, though its broader impacts on student confidence require further investigation across diverse educational contexts.

Introduction

One of the critical components for effective learning in the classroom is active engagement from students [1]. Asking questions or challenging assumptions invokes discussion and deepens conceptual understanding and strengthens critical thinking [2]. Although engagement can benefit many, students hesitate to participate in class. Feelings of nervousness, fear of embarrassment, and concerns about appearing unintelligent can create a psychologically unsafe environment, discouraging students from asking questions or correcting misunderstandings [3]. This can result in missed learning opportunities and the persistence of misconceptions. Prior research has demonstrated that instructor vulnerability, including openly acknowledging mistakes and admitting uncertainty, is an effective mechanism for building trust in the classroom. Without such vulnerability, instructors risk projecting an ideal persona that signals to students that perfection is required for success, which can discourage participation and risk-taking [4].

To address the challenge of psychological safety in classroom participation, the Instructor Mistake Meter (IMM) was created to encourage student participation and foster a psychologically safe classroom environment. The structure of the IMM is as follows: for every set number of instructor mistakes that the class identified and corrected during the lecture period, the entire class earned some amount of bonus points. For example, for every 10 mistakes made by the instructor, the entire class earned 1 bonus point at the end of the semester. The definition of a mistake was set at the start of the semester by the instructor. Some potential mistake types included mathematical, coding, spelling, and diagram drawing errors.

The IMM was developed and utilized for the first time in the Fall 2024 semester in an undergraduate core engineering class at Northeastern University, so its efficacy and impact on classroom dynamics have not yet been formally studied. Given the potential of this tool to improve classroom engagement, a systematic investigation of its effects was warranted. The objective of this study was to examine the effect of the IMM on undergraduate engineering student participation, academic confidence, and psychological safety in the classroom across two cohorts of students in the 2025 academic year.

Methods

The effect of the IMM was studied over two semesters in the same core undergraduate biomechanics class. The 100-minute class was typically split between 60 minutes of traditional lecturing and 40 minutes of active learning activities. The experiment semester in Fall 2025, which had 67 students split into two sections of 40 and 27 students, utilized the IMM. The control semester in Spring 2025, which had 62 students in one section, did not use the IMM but instead had an extra-credit assignment where students could earn as many bonus points as they would have from the IMM. For the bonus assignment, student teams developed their own exam based on the course content, along with a complete answer key. They were given two submission opportunities: the first to receive feedback, and the second as their final submission. The questionnaire and methodology of this study was approved by the Northeastern University Internal Review Board, approval number 24-07-30.

Instructor mistakes that were accounted for included arithmetic errors, coding syntax errors, missing components in free body diagrams, forgetting to include units, missing terms in equations, and using incorrect sign-notation in equations. Instructor mistakes were all unintentional mistakes made during the lecture. Both the control and experiment semesters used a Just-In-Time teaching approach where example problems were often generated the morning before or during class based on student feedback. These problems were solved in real-time in front of the class without a premade answer key which resulted in genuine instructor mistakes being made. In addition, many of the problems solved in class used MATLAB as a solving tool which provided students with opportunities to find coding syntax errors. Data was not collected on the ratio of mistakes made and corrected by the instructor compared to mistakes pointed out by students.

Data was collected from students in both semesters through optional anonymous surveys delivered in the first and last week of class. The surveys assessed student academic confidence and comfort in the classroom through Likert-Scale questions (Table 1). For the experiment semester, one additional short answer question was asked in the last week survey asking students to reflect on their experience with the IMM.

Table 1. Survey questions for first and last week surveys. Questions 1-6 were prefaced with "Answer the following questions on a scale from 1-5 where 1 is strongly disagree and 5 is strongly agree:". Question 7 was only asked to the experiment semester in the last week survey.

1) ...I am confident in my ability to understand complex course material.
2) ... I am comfortable asking questions in class.
3) ... I am confident in my ability to solve engineering problems.
4) ... I am confident in identifying others engineering mistakes.
5) ... I believe in my ability to overcome challenges.
6) ... I feel engaged during this class. - This was only asked in the last week survey
7) In your own words, how was your experience with the Mistake Meter tool in class?

In addition to the surveys, the total number of questions asked to the instructor was tracked over the course of the semester by a teaching assistant who attended the lecture. There were various qualifications for what counted as a question in this study. The question must have been asked by a student while the instructor was lecturing. If the instructor was working individually with a student during the active learning portion of the class, those questions were not counted. The question must have been about the content of the class. Questions about scheduling, deadlines, and other administrative matters were not counted.

The total number of questions asked was normalized by the number of students enrolled in each section across each day: experiment section 1, experiment section 2, experiment 1+2 sections, and control section. For the sake of analysis, experiment sections 1 and 2 were combined into experiment 1+2 section where the total number of questions each day was summed and divided by the sum of enrolled students in both sections. The number of questions per student was calculated across each lecture day and examined with a one-way ANOVA test ($p \leq 0.05$).

The Likert-Scale responses from the survey were examined using a 2-Factor ANOVA ($p \leq 0.05$) to analyze differences in responses between first and last week surveys as well as between the control and experiment semesters. Experiment data was combined when compared to the control data. Individually linked longitudinal analysis was not completed in this study. The anonymous nature of the survey and survey taking tool made it impossible to compare how individual students' answers changed between the first and last weeks of class. Question 6 assessing engagement was analyzed with a Wilcoxon rank sum test.

Open-ended survey responses were analyzed using thematic analysis as described by Braun and Clarke [5]. Thematic analysis was chosen since it is a widely utilized qualitative research tool and offered a structured yet flexible framework for identifying, analyzing, and interpreting patterns of meaning in a data set [6]. Furthermore, it enabled exploration of participant responses without requiring large sample sizes or extensive interview transcripts, making it more suitable for short open-ended survey data. Consistent with reflexive thematic analysis, meaning was interpreted through an iterative analytic process where themes were generated through both participants' responses and the research team's interpretation.

The qualitative analysis followed the six phases seen in Braun and Clarke's thematic analysis, starting with familiarization of the data. All survey responses were read multiple times to achieve a better understanding of the data. Initial ideas were recorded, and repeated words, ideas, and patterns were tracked. After identifying patterns, code words were created via an inductive open-coding approach. This allowed codes to emerge directly from the data rather than applying predetermined categories to reduce bias. Short descriptive texts were given to the data responses to capture their overall meaning. Some examples were sustained attentiveness, increased comfort in asking questions, and error-based learning. Some responses received multiple codes when more than one idea was expressed. Once the codes were completed and reviewed, they were grouped in broader themes that represent shared patterns of meaning. Once all the codes were compiled into preliminary themes, the themes were reviewed for internal coherence and distinction from one another. Any overlapping themes were examined, and codes were reassigned where necessary to ensure clarity and consistency. Lastly, themes were updated and clearly defined to capture their core meaning. Theme frequencies were then calculated by dividing the total appearances of the themes in each response by the total number of responses. Due to the open-ended survey's concise answers, themes were reported only at a single level and not subdivided into subthemes.

Results & Discussion

A total of 129 students were enrolled in this study. The control semester had 62 students while the experiment semester had 67 students split across two class sections with 41 and 27 students respectively. Not all students volunteered to take the first and last week surveys, and across both semesters survey response rates substantially decreased between the first and last week. The control semester collected 57 first week surveys and 34 last week surveys. The experiment semester collected 49 first week surveys and 27 last week surveys.

The analysis of student question-asking behavior revealed substantial differences between the experimental and control semesters. Through the semester, the experiment semester students asked 263 questions in section 1 and 187 questions in section 2. In comparison, the control semester students asked 176 questions (Table A1). Students in the experimental semester demonstrated significantly higher daily question rates compared to the control group (Figure 1). When data from both experimental class sections were combined, the average daily question rate per student was 0.339, representing a 139% increase over the control group's average of 0.142 questions per student each day. One-way ANOVA results confirmed statistically significant differences between all experimental sections and the control section. Experiment section 1 showed significantly higher questioning rates than the control group ($p = 0.008$), while experiment section 2 demonstrated an even stronger effect ($p < 0.001$). The combined experimental group, experiment 1+2, also differed significantly from the control ($p = 0.001$). No significant differences were observed between the experimental conditions themselves, indicating consistency in the intervention's effectiveness across both experimental classes. All question counter data were reported in the appendix (Tables A1 and A2).

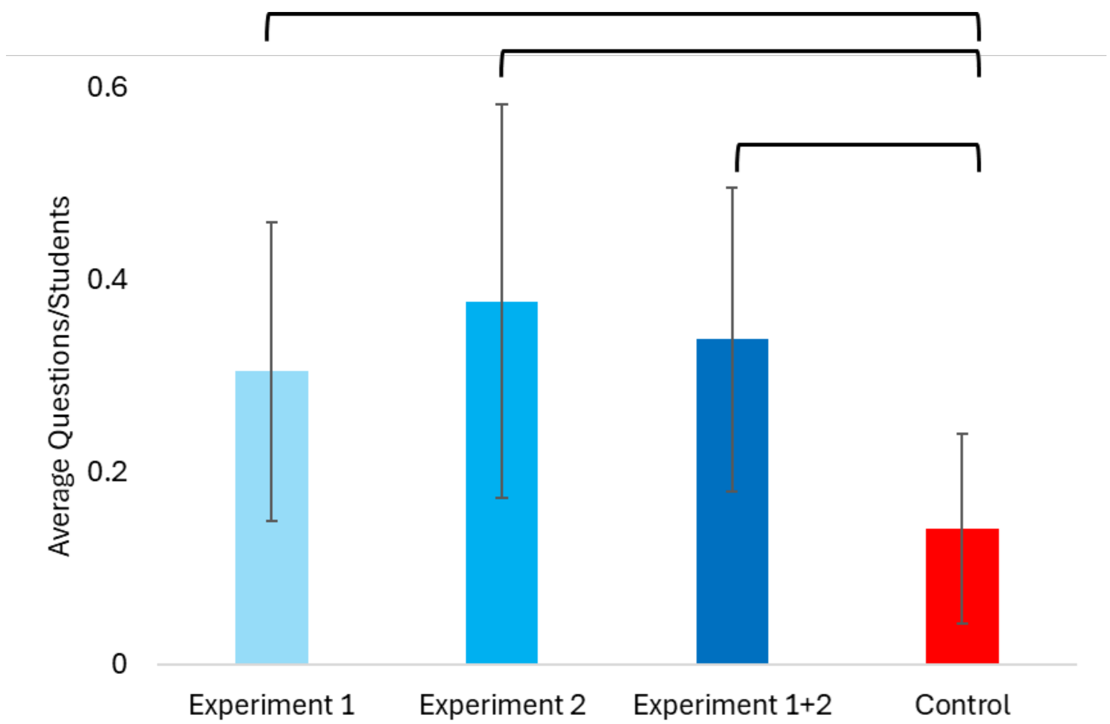

Figure 1. Average number of questions asked per student across each day of lecture. Error bars represent one standard deviation of the group's data. Brackets signify a statistically significant difference between groups ($p \leq 0.05$).

The average number of questions asked per student each class varied considerably across the semesters for all groups, yet distinct trends emerged in the experimental and control semesters (Figure 2). Second order polynomials were fit to the data to demonstrate trends (Table A1). The experimental sections reported a decrease in question asking mid-way through the semester followed by a strong increase toward the end of the semester, ending with more questions per student than at the start of the semester. In contrast, the control section exhibited a negative trend throughout the entire semester, ending with fewer questions per student asked than at the start of the semester. This downward trend was unexpected, as increased familiarity with the classroom environment, instructor, and peers was expected to foster greater psychological safety and, consequently, more questioning. The divergence between the experimental and control semesters suggested that the IMM played a meaningful role in encouraging student questioning behavior.

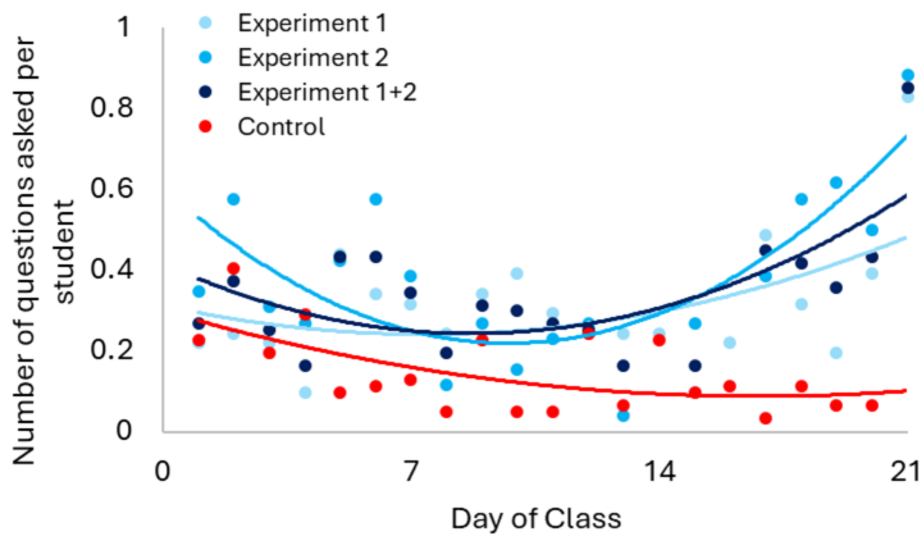

Figure 2. Number of questions asked per student each day of class and fit polynomial trendlines.

The qualitative short answer responses of the experiment semester students reported primarily positive impacts of the IMM. There were six themes that were identified in the short answer responses from the experiment semester (Table 2). Student engagement was the most frequent theme reported in the qualitative student survey responses, appearing in a total of 14 out of 26 answers. Experimental students described an increase in attentiveness, sustained focus, and greater awareness of potential mistakes during lectures. Students emphasized that the tool encouraged active review of the course material rather than passive listening. One response stated, “I think it helped people focus more and look for mistakes”, while another asserted, “Overall very positive, kept me attentive throughout the class lectures.” Others highlighted increased participation, explaining that the mistake meter encouraged them to remain vigilant and mentally engaged throughout class. These responses suggest that the IMM promoted class engagement and attentiveness during class. Full short answers and tracked themes were presented in the appendix (Tables A3, A4, and A5)

Table 2. Experiment semester short answer themes, definitions, and frequency of responses to the question "In your own words, how was your experience with the Mistake Meter tool in class?"

Theme	**Description**	**Number of responses with theme (n=26)**
Engagement	Increased focus, attentiveness, participation	14
Motivation	Fun, rewards, incentives	12
Learning	Clarification, conceptual understanding	9
Psychological safety	Comfort asking questions, reduced fear	8
Limitations	Trivial errors, distraction	3
Neutral	Minimal perceived impact	1

The second most prevalent theme revealed from the data was motivation. The motivation theme emerged from 12 responses, with many students referencing the point-based incentive of the tool. Several responses indicated that the opportunity to gain points for their final grade encouraged them to pay closer attention and remain involved in class. A student commented, "I think it was a good motivator," with another stating, "It made me pay more attention to details in class and gave me a little buffer for my grade." Further responses described the activity as enjoyable and rewarding, contributing to a positive classroom atmosphere.

Learning and Conceptual understanding appeared throughout 9 responses. Students noted that identifying mistakes facilitated clarification and reinforced understanding. Many of the responses reported the importance of learning from errors. One commented, "Even when it wasn't truly a mistake, it allowed the professor to re-explain that part." Another noted, "It was a good way to learn from someone else's mistakes and avoid making them in the future." Some students also described improved retention of the material when mistakes were discussed and corrected. These responses suggest that the mistake meter supported error-based learning, reinforcing a better understanding of the material.

Another theme generated was psychological safety, appearing in 8 out of the 26 responses. Students reported feeling more comfortable asking questions and speaking up to identifying potential mistakes without embarrassment or judgment. Several participants noted that seeing the instructor openly acknowledging mistakes normalized them and reduced students' fear in making their own. One response reported, "By the end of the semester I felt very comfortable asking questions in class," and another student explained that the tool "was something that occasionally brought excitement or an overall safe, warm dynamic in the classroom knowing the professor is being accounted for on his work in class the same way we are in ours outside of class." The remaining responses highlighted the tool's ability to create a supportive and open class environment.

Although most responses were positive, there were some students that felt the tool was distracting. Three of the responses revealed some concerns, mainly that some mistakes seemed trivial or unrelated to the course concepts. One participant remarked, "Students were just trying to catch the professor in their random errors that never really had to do with material." Another student expressed that some errors were overly technical or minor. Although such concerns were relatively infrequent, these responses revealed some potential limitations of the tool. In addition, one student response reflected a neutral perception of the Mistake Meter, indicating a minimal impact on the student's experience of the course. The student stated, "…most of the time, I did not pay much attention to it."

Overall, the qualitative findings indicate that the IMM had a mostly positive impact on most students through increasing engagement and motivation. A smaller but substantial number of students reported learning-related improvements and more comfort asking questions. These results suggest that the tool positively affects multiple aspects of the classroom experience for most participants. It should be noted that these findings are based on a limited number of short open-ended responses, which may constrain the depth of qualitative interpretation.

Quantitative analysis of student academic confidence and comfort in the classroom did not find a significant difference between the first and last week for both control and experimental semesters, except for the control group's confidence in solving engineering problems (Figure 3). For this analysis, the experiment semester sections were combined and compared to the control semester. There was a significantly higher confidence score in the last week of the control semester compared to the first week of the experimental semester for understanding complex course material ($p = 0.02$), comfortability asking questions ($p = 0.002$), solving engineering problems ($p < 0.001$), and identifying mistakes ($p < 0.001$). The time between first and last week of each study had a significant impact on confidence in understanding complex course material ($p = 0.002$), comfortability asking questions ($p = 0.006$), confidence in solving engineering problems ($p = 0.017$), and identifying mistakes ($p = 0.002$) (Table A6 and A7).

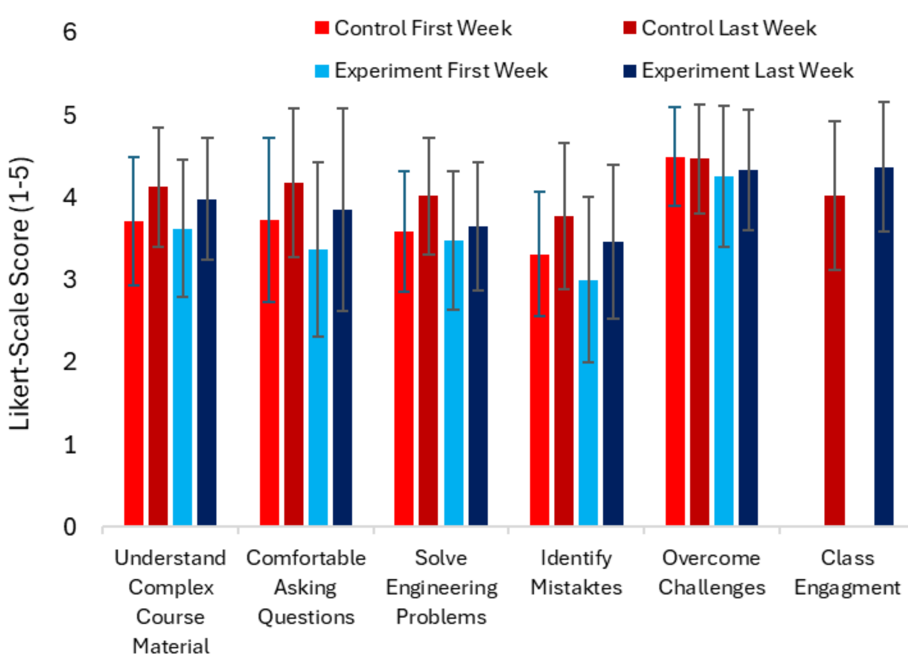

Figure 3: Average Likert-scale scores assessing confidence in academic confidence and comfort. Error bars represent one standard deviation.

While not significant, changes in survey responses to questions related to classroom engagement were higher in the experiment semester compared to the control semester. There was a 14% increase in confidence towards asking questions in the class, compared to a 12% increase in the control group. The confidence score in the experimental semester was 0.37 higher than that of the control group in response to feeling engaged in the classroom ($p = 0.09$).

As survey participation was optional, not all students in the study responded, and there was a decrease in responses in both control and experimental semesters for the last week. Further, with anonymous responses, it wasn't possible to track changes in individual students' scores. The results of the quantitative anonymous survey suggest that the IMM did not have a substantial effect on student confidence or comfort in the classroom. Additionally, the attrition rate of the survey responses reduces the impact of these results as nearly half of each class failed to answer. The low response rates in the last week surveys may have introduced non-response bias, as students who chose to respond may not be representative of the full class population. Future iterations of this study should administer surveys in class rather than online and offer incentive to complete them to reduce attrition.

There were three important limitations to this study, section size, curriculum changes, and team dynamic changes. This study was run across two different semesters of the same course, but the control semester was one large class section while the experiment semester was split into two smaller sections. The control section had more students to potentially ask more questions each class, but the larger class size may have decreased comfort in asking questions in comparison to the smaller experiment sections. Matching the section structure of the classes would have improved the experimental set up, but this was outside the control of the research team.

This study was run across two different semesters of the same core undergraduate engineering course, Biomechanics. While the course title and instructor remained the same across the two semesters, there were substantial updates made to the curriculum and classroom approach that may have affected the results of this study. The control semester curriculum covered deformable solids and rigid body dynamics. In comparison, the experiment semester curriculum removed rigid body dynamics, expanded on deformable solids, and added finite element analysis. The classroom topics were changed but the level of difficulty was meant to remain constant across both semesters. This change was outside the control of the research team, but future work should examine courses with fewer curricular changes to strengthen the validity of the study.

The final limitation was student team structure changes. In the control semester, students were assigned to teams of four where they would be asked to complete their homework together and work in class on problems together. The control semester teams were implemented to increase classroom comfort by manufacturing study groups for students. This team aspect of the class was discontinued for the experiment semester because of the negative feedback from the control students who reported challenges navigating team logistics during the course. The effect of these student teams was unknown, but they may have influenced classroom psychological safety and question asking behavior.

The results of this study suggest the IMM had primarily positive or neutral impacts on students. This low-effort high-yield tool may be a useful tool to implement into other classrooms, but before large suggestions are made on IMM implementation, it is important to acknowledge the scope and limitations of this study. The IMM has only been implemented in a single class, by a single instructor, in a single institution. Different course structures that rely less on problem solving in class may need to restructure the IMM to be more effective for their structures. In addition, the confounding factors of curriculum, team structure, and section sizes all had an unknown effect on the students. Future work should ensure the control and experiment semesters are better aligned to strengthen the impact of the study.

It is important to note that all of the instructor's mistakes in this study were unintentional. The level of the instructor, junior versus senior, may have an impact on the IMM's effectiveness and effort to implement. More senior instructors may not make as many mistakes and would need to intentionally add errors to their lectures for the IMM to be effective. In this case, the IMM would require additional effort to be implemented. The effect of intentional errors has not been studied at the undergraduate level and there may be an optimal range of intentional errors to add to improve outcomes. Before intentional mistakes are to be used in the IMM, research should be completed examining the appropriate number of mistakes made per class to maximize the positive effects of this tool. In addition, further research on the IMM should be completed across different classes, instructors, and institutions before wide implementation suggestions are made.

Conclusion

The objective of this study was to assess the effect of the IMM on student confidence, comfort, and level of classroom engagement. Quantitative survey results reported that the IMM did not have a significant impact on student confidence and comfort in the classroom. Thematic analysis of the qualitative student reflections reported mostly positive experiences with the IMM as a tool to increase student comfort. The most frequent themes identified from participants' responses, student engagement and motivation, suggest that the tool encouraged better attentiveness, active review of material, and increased participation during class. A substantial number of responses recorded other learning benefits as well, including clarification of confusing concepts and reinforcement of understanding of course material. Several students highlighted an increased comfort in asking questions and reduced fear of speaking up, indicating improvements in psychological safety within the classroom environment. In addition to the positive qualitative responses, the experiment semester asked more than twice the number of questions per student than the control semester, suggesting an increase in comfort asking questions. Overall, the IMM reported either neutral or positive effects on student confidence, comfort, and classroom participation. While successful, before implementation recommendations are made, the tool should be tested with other classes, instructors, and institutions.

References

[1] R. Strachan and L. Liyanage, "Active Student Engagement: The Heart of Effective Learning," in *Global Innovation of Teaching and Learning in Higher Education: Transgressing Boundaries*, P. C. Layne and P. Lake, Eds., Cham: Springer International Publishing, 2015, pp. 255–274. doi: 10.1007/978-3-319-10482-9_16.

[2] T. Tofade, J. Elsner, and S. T. Haines, "Best Practice Strategies for Effective Use of Questions as a Teaching Tool," *Am. J. Pharm. Educ.*, vol. 77, no. 7, p. 155, Sep. 2013, doi: 10.5688/ajpe777155.

[3] E. M. Nadile *et al.*, "Call on me! Undergraduates' perceptions of voluntarily asking and answering questions in front of large-enrollment science classes," *PloS One*, vol. 16, no. 1, p. e0243731, 2021, doi: 10.1371/journal.pone.0243731.

[4] S. Huddy, "Vulnerability in the Classroom: Instructor's Ability to Build Trust Impacts the Student's Learning Experience," *Int. J. Educ. Res.*, vol. 10, no. 2, pp. 96–103, 2015.
[5] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006, doi: 10.1191/1478088706qp063oa.
[6] S. K. Ahmed *et al.*, “Using thematic analysis in qualitative research,” *J. Med. Surg. Public Health*, vol. 6, p. 100198, Aug. 2025, doi: 10.1016/j.glmedi.2025.100198.

Appendix

Table A1. Number of questions asked across each day of class for each class section and polynomial equations fit to data

Number of Questions Asked Each Class				
Day of Class	**Experiment 1**	**Experiment 2**	**Experiment 1&2**	**Control**
1	9	9	18	14
2	10	15	25	25
3	9	8	17	12
4	4	7	11	18
5	18	11	29	6
6	14	15	29	7
7	13	10	23	8
8	10	3	13	3
9	14	7	21	14
10	16	4	20	3
11	12	6	18	3
12	10	7	17	15
13	10	1	11	4
14	10	N/A	N/A	14
15	4	7	11	6
16	9	N/A	N/A	7
17	20	10	30	2
18	13	15	28	7
19	8	16	24	4
20	16	13	29	4
21	34	23	57	N/A
Total Questions	**263**	**187**	**430**	**176**
Polynomial Fit Equation	$y = 0.0013x^2 - 0.0193x + 0.3124$	$y = 0.0041x^2 - 0.0794x + 0.6061$	$y = 0.0023x^2 - 0.0398x + 0.4165$	$y = 0.0008x^2 - 0.0252x + 0.2989$
R^2 **Coefficient**	0.22	0.585	0.443	0.369

Table A2. Single Factor ANOVA p-values for questions per students across all four groupings. Experiment 1+2 data excluded days 14 and 16 because Experiment section 2 lacked data.

Cohort		**p-value**
Experiment 1	Experiment 2	0.467
Experiment 1	Experiment 1+2	0.911
Experiment 1	Control	0.008
Experiment 2	Experiment 1+2	0.864
Experiment 2	Control	< 0.001
Experiment 1+2	Control	0.001

Table A3. Theme key

Theme	**Theme key**
Engagement and Attention	T1
Psychological Safety	T2
Learning Outcomes	T3
Motivation and Enjoyment	T4
Limitations	T5
Neutral	T6

Table A4. Short qualitative survey responses of experiment semester last week survey question "In your own words, how was your experience with the Mistake Meter tool in class?"

Raw Response	**Code**	**Theme(s)**
I think it was a good motivator.	Reward-driven attention, extrinsic motivation	T4,
I really enjoyed the mistake meter. I think it encourages people to stay engaged and I also noticed people asked a lot more questions than they probably would have. I would recommend the mistake meter for any class	Increased engagement, Increased comfort asking questions	T1, T2
I think it helped people focus more and look for mistakes	Sustained attentiveness	T1
I liked having the mistake meter in class. It made me feel more engaged in class and by the end of the	Increased engagement, Increased comfort asking questions, extrinsic motivation	T1, T2, T4

semester I felt very comfortable asking questions in class.		
I liked it, it was a good motivator to get students to pay attention and understand what's being written in the notes, and it is beneficial to our grade.	Extrinsic motivation, Sustained attentiveness, improved conceptual understanding,	T1, T3, T4
The mistake meter was helpful because it encouraged students to speak up whenever something didn’t make sense. Even when the issue wasn’t truly a mistake, students felt more comfortable questioning computations that seemed incorrect, which led to either a clarification or the confirmation that an error had been made. This prevented them from memorizing inaccurate information or from stopping paying attention to the lesson due to confusion.	Increased comfort asking questions, facilitated clarification, prevented misconceptions,	T2, T3,
Very fun. It incentivized me to focus more in class to "catch mistakes" and helped me feel more comfortable in calling out if I thought the professor made a mistake. Even if it wasn't a mistake, if someone pointed out something that looked confusing, it allowed the professor to re-explain that part.	increased vigilance for errors, increased comfort asking questions, sustained attentiveness, facilitated clarification	T1, T2, T3
It was occasionally nice to see the mistake meter go up, but most of the time, I did not pay much attention to it. I don't think it impacted my experience in this class in a meaningful way.	minimal perceived impact,	T6

I kind of thought it was weird. Students were just trying to catch the professor in their random errors that never really had to do with material.	low relevance of errors, Trivial error focus,	T5
It was good, it helped show how easy it is to make mistakes and it helps when it comes to learning	normalized instructor fallibility, error-based learning	T2, T3,
It made me pay more attention to details in class and gave me a little buffer for my grade (which I need)	Sustained attentiveness, extrinsic motivation,	T1, T4
I think it was good. Sometimes it made me feel bad pointing them out if there were a lot of mistakes, but I would have felt worse if there was not a mistake meter, so for that reason it was good. Took some of the pressure off of the professor being annoyed for pointing out the mistake.	lowered social risk, normalized instructor fallibility,	T2, T3
Overall very positive, kept me attentive throughout the class lectures.	Sustained attentiveness,	T1
It was good but some were a little ridiculous to count like not having a coordinate system, forgetting a negative sign, typos, and things like that. It was good when it was conceptual ones, mistakes in the homework answer key, mistakes in the printed exam and things like that. I feel like it kept class active and motivated paying close attention so that was so good!	increased engagement, active participation, improved conceptual understanding, Trivial error focus, extrinsic motivation	T1, T3, T4, T5
Great	general positive perception	T4

My experience was overall good- it was something that occasionally brought excitement or an overall safe, warm dynamic in the classroom knowing the professor is being accounted for on his work in class the same way we are in ours outside of class. It didn't make any significant change in the course in my opinion, and sometimes got a little too technical/picky from the students, but overall, I had a good experience and it brought more good than bad to the course.	lowered social risk, normalized instructor fallibility, trivial error focus	T2, T5
It is great	general positive perception	T4
Made me want to pay attention more.	Sustained attentiveness,	T1
I liked it because it was a good way to learn from someone else's mistakes and it allowed me to see errors that I could potentially make, helping me to avoid them in the future.	increased vigilance for errors, error-based learning,	T1, T3
Good i think it is really useful and logical	general positive perception	T4
I think that having the incentive of the extra credit made me more engaged in class so that I would know if there was a mistake. I also think it was nice for a prof to be open about making mistakes and not trying to come across like he knows everything.	extrinsic motivation, increased engagement, normalized instructor fallibility,	T1, T2, T4
I liked it, I feel like I was really thinking critically about the work being done in class rather than passively listening. I think it also made	increased engagement, critical thinking engagement, knowledge retention,	T1, T3

me retain the information where a mistake was made.		
Pretty good, encouraged me to pay close attention and to make sure i understood what I was doing so I could help my class get those extra points	sustained attentiveness, prevented misconceptions, extrinsic motivation	T1, T3, T4
It was fun because it felt rewarding and it made the class feel more approachable and doable.	gamified engagement, prevented misconceptions, extrinsic motivation,	T1,T4
It sometimes made class more fun	Reward-driven attention	T4

Table A5. Codebook for themes tracked in short responses

Student Engagement & Attention Theme Codes				
Code	**Definition**	**Include When**	**Exclude When**	**Example Quote**
Increased engagement	Student reports being more involved or mentally engaged during class	Mentions engagement or participation	Only motivation/enjoyment	“It encourages people to stay engaged.”
Sustained attentiveness	Student reports increased focus, attention, or concentration during lecture	Mentions focus or alertness	Only enjoyment	“It helped people focus more.”
Active participation	Student describes participating more actively or contributing verbally	Mentions speaking or contributing	Passive observation	“It kept class active.”
Increased vigilance for errors	Student actively monitors material for mistakes	Mentions catching mistakes	General focus only	“Look for mistakes.”
Critical thinking engagement	Deeper analytical thinking	Mentions thinking critically	Simple attention	“Thinking critically rather than passively listening.”
Psychological Safety & Participation Confidence Theme Codes				

Code	Definition	Include When	Exclude When	Example Quote
Increased comfort asking questions	Student feels more comfortable speaking up or asking questions.	Mentions comfort or confidence	General participation only	“Felt comfortable asking questions.”
Lowered social risk	Student perceives reduced fear of judgment or embarrassment.	Mentions reduced pressure or fear	Instructor behavior only	“Took some of the pressure off.”
Normalized instructor fallibility	Instructor mistakes viewed as normal	Mentions openness to mistakes	Mentions errors only	“Nice for a prof to be open about making mistakes.”
Learning & Conceptual Understanding Theme Codes				
Code	**Definition**	**Include When**	**Exclude When**	**Example Quote**
Improved conceptual understanding	Better comprehension of material	Mentions understanding	Only motivation	“Understand what’s written in the notes.”
Facilitated clarification	Mistakes prompt clarification	Mentions re-explanation	Understanding without explanation	“Allowed the professor to re-explain.”
Error-based learning	Learning from mistakes	Mentions learning from errors	Errors without learning outcome	“Learn from someone else’s mistakes.”
Prevented misconceptions	Avoided incorrect learning	Mentions preventing confusion	General learning only	“Prevented memorizing inaccurate information.”
Knowledge retention	Improved memory of material	Mentions retaining or remembering	Short-term understanding	“Made me retain the information.”
Motivation & Enjoyment Theme Codes				
Code	**Definition**	**Include When**	**Exclude When**	**Example Quote**
Extrinsic motivation	Motivation driven by grades or points	Mentions credit or grade	Fun only	“Buffer for my grade.”
Reward-driven attention	Rewards increase focus	Mentions motivation due to reward	Enjoyment only	“Good motivator.”

Increased enjoyment	Positive emotional response	Mentions fun or enjoyable	Motivation only	“Made class more fun.”
General positive perception	Non-specific positive evaluation	Short praise statements	Specific outcome mentioned	“Great.”
Perceived Limitations & Distractions Theme Codes				
Code	**Definition**	**Include When**	**Exclude When**	**Example Quote**
Low relevance of errors	Errors not related to learning goals	Mentions irrelevant errors	Minor error complaints	“Random errors not related to material.”
Trivial error focus	Focus on insignificant errors	Mentions typos or small mistakes	Relevance issues	“Forgetting a negative sign.”
Over-technical nitpicking	Excessive technical focus	Mentions overly picky behavior	Triviality only	“Too technical/picky.”
Neutral / Minimal Impact Theme Codes				
Code	**Definition**	**Include When**	**Exclude When**	**Example Quote**
Minimal perceived impact	Little or no effect reported	Mentions no meaningful impact	Positive or negative judgment	“Did not impact my experience.”

Table A6. Results of 2-way Anova of Survey Questionnaire

Question 1	**P-value**
Study	0.360
Time	0.002
Study:Time	0.83
Question 2	**P-value**
Study	0.042
Time	0.006
Study:Time	0.914
Question 3	**P-value**
Study	0.057
Time	0.017
Study:Time	0.298
Question 4	**P-value**
Study	0.033
Time	0.002
Study:Time	0.990
Question 5	**P-value**
Study	0.109
Time	0.804
Study:Time	0.670

Table A7. Post-hoc Analysis of 2-way Anova of Survey Questionnaire

Question 1		**P-value**
Control First	Experiment First	0.938
Control First	Control Last	0.068
Control First	Experiment Last	0.442
Experiment First	Control Last	0.020
Experiment First	Experiment Last	0.217
Control Last	Experiment Last	0.894
Question 2		**P-value**
Control First	Experiment First	0.279
Control First	Control Last	0.189
Control First	Experiment Last	0.956
Experiment First	Control Last	0.003
Experiment First	Experiment Last	0.206
Control Last	Experiment Last	0.617
Question 3		**P-value**
Control First	Experiment First	0.887
Control First	Control Last	0.050
Control First	Experiment Last	0.987
Experiment First	Control Last	0.010
Experiment First	Experiment Last	0.795
Control Last	Experiment Last	0.247
Question 4		**P-value**
Control First	Experiment First	0.286
Control First	Control Last	0.073
Control First	Experiment Last	0.876
Experiment First	Control Last	< 0.001
Experiment First	Experiment Last	0.1303
Control Last	Experiment Last	0.5269
Question 5		**P-value**
Control First	Experiment First	0.329
Control First	Control Last	0.999
Control First	Experiment Last	0.782
Experiment First	Control Last	0.533
Experiment First	Experiment Last	0.969
Control Last	Experiment Last	0.880