WIP: Reflection Trends in Sequential First-Year Engineering Courses

Mr. Alfa Satya Putra, Purdue University – West Lafayette (College of Engineering)

Alfa Satya Putra is a PhD candidate at School of Engineering Education at Purdue University. He has Bachelor's and Master's degree in Electrical and Computer Engineering from Purdue University. Before joining the PhD program, Alfa has served as a lecturer in Indonesia. Alfa is mainly interested in investigating the implementation of reflective activities in large classrooms and assessing how reflective activities affect student learning and academic performance.

Dr. Muhsin Menekse, Purdue University – West Lafayette (College of Engineering)

Muhsin Menekse is an Associate Professor at Purdue University with a joint appointment in the School of Engineering Education and the Department of Curriculum & Instruction. Dr. Menekse's primary research focuses on exploring K-16 students' engagement and learning of engineering and science concepts by creating innovative instructional resources and conducting interdisciplinary quasi-experimental research studies in and out of classroom environments. Dr. Menekse is the recipient of the 2014 William Elgin Wickenden Award by the American Society for Engineering Education. He is also selected as an NSF SIARM fellow for the advanced research methods for STEM education research. Dr. Menekse received four Seed-for-Success Awards (in 2017, 2018, 2019, and 2021) from Purdue University's Excellence in Research Awards programs in recognition of obtaining four external grants of $1 million or more during each year. His research has been generously funded by grants from the Institute of Education Sciences (IES), the U.S. Department of Defense (DoD), Purdue Research Foundation (PRF), and the National Science Foundation (NSF).

Muyang Wu, Southern Methodist University (Department of Computer Science & Engineering)

Muyang Wu is a PhD student in Computer Science at Southern Methodist University. Her research focuses on human-computer interaction and AI in education.

WIP: Reflection Trends in Sequential First-Year Engineering Courses

Abstract

[1]WIP: This Work in Progress Paper explores how reflection trends vary across two sequential first-year engineering (FYE) courses at a large Midwest Public University. While prior studies found association between reflective activities in undergraduate STEM classrooms and learning outcomes, investigating reflection trends between sequential FYE courses remains a relatively underexplored topic. In this study, we analyzed student reflections on what they found confusing and interesting from two sequential FYE courses, referred to as ENGR 101 (N=80) and ENGR 102 (N=68). We used a mobile application to prompt students to reflect on what they found confusing and interesting after every lecture during two academic semesters. We conducted text analysis to uncover and analyze emerging patterns in student reflections. We created an initial codebook using Chi and VanLehn's physics self-explanation framework and modified our codebook along the coding process. Two coders conducted independent coding on a subset of reflections and calculated inter-rater reliability using Cohen's κ. After establishing reliability, the validated codebook was applied to a larger reflection dataset using human coding alongside large language model (LLM) coding. We quantified code frequencies and compared their distributions across courses and reflection types.

Our preliminary findings indicate that ENGR 101 reflections exhibit a more balanced mix of conceptual and procedural content, with frequent mentions of statistics and Excel. In contrast, ENGR 102 reflections are dominated by technical procedures related to MATLAB programming and file management. Confusing reflections highlight conceptual hurdles in ENGR 101 and procedural challenges in ENGR 102, supporting prior observations that programming courses elicit more procedural confusion. These patterns suggest that reflection content systematically differs between sequential FYE courses, offering insight into how instructors may better scaffold the transition from statistical reasoning to computational problem-solving.

Introduction

Reflection in education has gained significant attention in STEM education in recent years [1]. Multiple studies have explored integrating reflective activities in STEM courses and assessed their impact on student learning metrics. A meta-analysis by [2] found that reflection interventions have medium-sized effect ($g = 0.56$, $SE = 0.06$) on learning outcomes. Another meta-analysis by [3] found that digital tools to foster self-reflections have positive effect on academic achievement ($d = 0.42$), self-regulated learning ($d = 0.19$), and motivation ($d = 0.17$).

In this study, we implemented reflective activities in the form of technology-assisted scaffolding to help students reflect on what they found confusing and interesting from their learning. Multiple studies have explored how learning confusion and interest can reveal valuable information of students' learning experiences [4], [5]. In the context of FYE courses, prior studies have examined how students' reflection behaviors are associated with learning engagement and academic performance [6], [7]. However, investigating reflection trends

---

[1]This work was supported by NSF #2507035. The opinions expressed do not represent the views of the U.S. National Science Foundation.

between sequential FYE courses remains relatively underexplored. In this study, we investigate reflection trends in two FYE courses covering different sets of core engineering concepts. We hypothesize that analyzing student reflections can reveal patterns of learning confusion and interest. These patterns can be utilized by educators and researchers to design learning interventions, inform strategies to scaffold course transitions, and ultimately contribute to improvements in student motivation, academic performance, and retention. We also aim to explore how Large Language Models (LLMs) can be leveraged to provide scalable support to analyze reflection patterns while maintaining human-in-the-loop to ensure effective human-AI collaboration in research.

Literature Review

FYE Course Progression

FYE courses are typically designed to introduce fundamental skills such as problem-solving, data analysis, engineering reasoning, and programming that are central to engineering practice [8], [9]. A well-designed sequence of FYE courses supports students' engineering thinking development and allows concepts introduced in one course to be reinforced in the next [10]. Large enrollments create challenges for maintaining engagement and providing individualized feedback [8], [11]. To address these challenges, effective FYE instruction emphasizes active learning, timely feedback, and reflective opportunities that allow students to monitor their understanding and learning progress [12]. Reflection provides valuable insight and supports instructors in identifying learning needs across sequential courses.

Reflection as a Learning Process

Reflection links learning with metacognition and self-regulated learning, using Schön's reflective practitioner as framework to view learning as interaction between experience and interpretation [13], [14]. When students encounter difficulty, reflection helps them examine their reasoning, identify knowledge gaps, revise their approach, and develop self-regulated learning [15]. Structured prompts and scaffolds can be utilized to support students to reflect meaningfully [16], [17]. Tools such as guiding questions and visual support help students connect new and prior knowledge [4], [18]. In large STEM courses, reflection also promotes engagement by allowing students to express confusion and highlight meaningful learning experiences [1].

Digital Tools for Reflection Support

Digital technologies are widely used to support reflection in various learning settings. Mobile platforms allow students to record reflections close in time to learning activities, increasing participation and scalability in large courses [19]. Many tools use structured prompts to support students to write specific reflections. For example, ReFlex and TeamUp use scaffolded prompts and peer interaction to support deeper reflection [20], while Reflect.UP uses notifications and competency-based questions to encourage ongoing self-assessment [21]. Other tools use natural language processing to identify shared challenges and misconceptions [22], [23].

Confusion and Interest Focused Reflection Prompts

Confusion and interest are important emotions that influence learning. Both arise when new information challenges existing understanding [24]. Confusion can support learning by signaling gaps in understanding and encouraging students to seek clarification. Interest reflects a positive response to challenges and supports attention and curiosity [25]. Reflection prompts focused on confusion and interest help students make these experiences explicit. Articulating confusion supports metacognitive awareness and self-regulation [26], while reflecting on interest reinforces motivation and perceived relevance. For instructors, these prompts provide insight into where students struggle or engage, offering a scalable way to monitor learning in large FYE courses.

LLM-Supported Analysis of Textual Data

Advances in large language models (LLMs) have expanded options for analyzing reflective writing on a scale. [27] showed that LLMs could classify reflection dimensions with moderate to high reliability (Cohen's $\kappa = 0.61$-0.85). However, they were limited in flexibility and contextual understanding, as LLMs analyze text using natural language prompts rather than predefined features. [28] found that models such as GPT-4 can perform qualitative coding with performance comparable to acceptable human agreement ($F1 \approx 0.8$). [29] reported that LLMs provide consistent results when paired with human review, supporting hybrid workflows where models handle scale and humans focus on interpretation. This paired human-AI approach supports efficient and transparent analysis of large reflection datasets.

Research Gap and Research Questions

Prior research on reflection in engineering education has focused on learning progress and metacognition, examining the relationship between reflection quality and performance. Fewer studies have compared reflection patterns across sequential FYE courses, where students transition from foundational topics such as statistics and project management to more advanced computational and programming tasks. This transition represents an important developmental phase for engineering learners, yet the evolution of their reflective practices across courses remains underexplored. Furthermore, earlier studies aggregate reflections across a semester, overlooking temporal changes in reflection. By examining reflections across different time periods in a semester, this study captures shifts in confusion and interest as courses progress. The study also introduces a hybrid human-LLM coding approach that combines interpretive judgment with scalable analysis.

We formulate our research questions as follows:

- RQ1: What topics or concepts are most frequently associated with confusion or interest in ENGR 101 and ENGR 102?
- RQ2: How do topics and concepts most frequently associated with confusion or interest overlap between ENGR 101 and ENGR 102?

Methodology

Participants and Data Collection

We collected reflections on learning confusion and interest from one course section of two sequential FYE courses in a large Midwest Public University, which we refer to as ENGR 101

(N=80) and ENGR 102 (N=102). ENGR 101 introduces students to foundational engineering skills, including systems approach, model development, quantitative analysis, and modern engineering tools such as Excel and MATLAB. ENGR 102 builds upon the skills developed in ENGR101 with emphasis on project management, engineering design, data analytics, and deeper engagement with modern engineering tools. Both courses emphasize the development of communication and teamwork skills and involve team-based assignments and projects, in addition to individual assessments. Both course sections were taught by the same instructor in two different semesters, and the authors were not involved in teaching both courses.

We used a mobile application called CourseMIRROR, which is designed to collect and summarize student reflections [30]. We obtained IRB approval for this study and were assigned protocol number MREB 6067. Students were prompted to write a reflection on what they perceive as confusing and a reflection on what they perceive as interesting after every lecture. Reflection writing was not mandatory, and the length of student reflections vary between one to three sentences. Reflections were collected from 19 lectures in ENGR 101 and 24 lectures in ENGR 102. The difference in number of lectures is due to variations in course structure and assignments between both courses.

Data analysis

We implemented inductive and deductive coding to categorize students' self-reflections in order to gain a better understanding of students' learning experiences [31], [32]. We used Chi and VanLehn's physics self-explanation framework, which categorized constituent knowledge into four different categories [33]. This framework has been used in a previous study to categorize reflection trends in an introductory physics course [34]. We consider this framework appropriate for this study because the categories in this framework align well with core engineering concepts taught in ENGR 101 and 102. However, during the preliminary stage of codebook development, we found that systems thinking category did not map well with student reflections from both courses. We removed this category and retained three categories: technical procedures, principles, and conceptual knowledge. Next, we conducted inductive coding by having two coders to independently code a subset of the reflection dataset. The coders identified three additional categories for reflections focusing on instructional context, learning metacognition, and a miscellaneous category for reflections that did not correspond with existing categories.

To examine how reflection trends change across a semester, we divided reflection data into three subsections, namely early, mid, and late, based on major milestones in each course. We consulted with the instructor to gain clearer understanding of course structure and timing of major assignments in each course. For ENGR 101, we define early subsection as reflections collected from weeks 1 to 5, mid from 6 to 9, and late from 10 to 16. For ENGR 102, we define early as reflections collected from weeks 1 to 6, mid from 7 to 11, and late from 12 to 16.

We conducted two rounds of reflection coding between two human coders and calculated inter-rater reliability between coders to ensure coding consistency and usability of the codebook [35]. For the first round of coding, each coder conducted independent coding using an established codebook on 40 reflections from the first week of confusion and interest reflections from both courses. We addressed our coding discrepancies and continued with the second round of coding, where we coded 240 confusion and interest reflections from the first week of the three subsections from both courses. Cohen's Kappa value for our second round of coding was κ = 0.884. Our final coding scheme consists of six codes, namely Technical Procedure (Code 1), Conceptual Knowledge (Code 2), Principle-Related Knowledge (Code 3), Instructional Context (Code 4), Metacognition (Code 5), and Miscellaneous (Code 6). Detailed explanation of each code can be seen in Appendix A.

Following the two rounds of human coding, the combined set of 240 human-coded reflections served as the foundation for hybrid human-LLM approach to apply the finalized coding scheme to the full reflection dataset. To support consistent and repeatable coding, we created a customized GPT called "Student Reflection Coding Assistant". The customized model was configured with fixed instructions, the finalized coding schema, and human-coded examples, allowing it to apply the same coding rules and output format across all batches. The LLM was instructed to assign one code to each reflection using the predefined categories to provide a brief rationale citing key phrases from the reflection text. Our instruction for the customized model can be seen in Appendix B, and sample inputs and outputs can be seen in Appendix C.

Reflections were coded in batches of 20, with contextual information provided for each batch, including course and prompt type. To evaluate alignment between human and LLM coding, we compared LLM-generated codes with human codes for 123 reflections from the first week of remaining data across both courses. Cohen's κ for this comparison was 0.905, indicating strong agreement. Following this validation step, the customized LLM was used to code the full dataset. As an ongoing quality check, a human coder randomly recoded approximately 20% of reflections within each batch and compared them with the LLM output. Agreement remained consistently high across all batches. The final coded dataset consisted of 1,885 reflections from ENGR 101 and 2,262 reflections from ENGR 102. These coded data were then used to quantify code frequencies and compare distributions across courses, reflection types, and time periods.

Preliminary Results and Discussion

Overall Distribution of Reflection Categories

Across both courses, the most frequently occurring categories were Technical Procedure and Instructional Context, indicating that students tend to focus on carrying out tasks and responding to course activities and instructional design. In contrast, Principle and Metacognition were the least frequent.

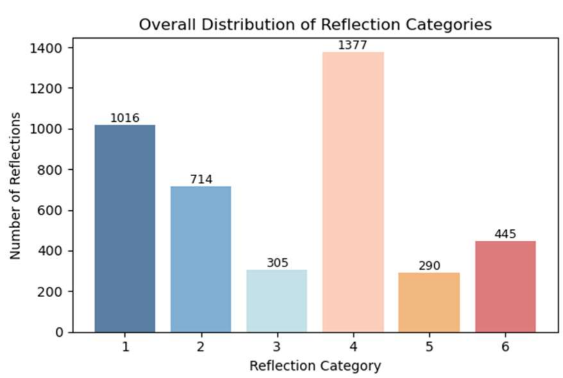

Figure 1. Overall Distribution of Reflection Categories

Confusion vs. Interest Patterns

Clear differences emerged between confusing and interesting reflections. Interesting reflections were more likely to be coded as Technical Procedure, Principle, Instructional Context, and Metacognition. These reflections often described engaging activities, successful task completion, reasoning about content, or awareness of learning progress. In contrast, confusing reflections showed higher proportions of Conceptual Knowledge and Miscellaneous. This suggests that confusion was more often associated with uncertainty about definitions or expressed in vague or unspecific ways, particularly when students struggled to articulate the source of difficulty.

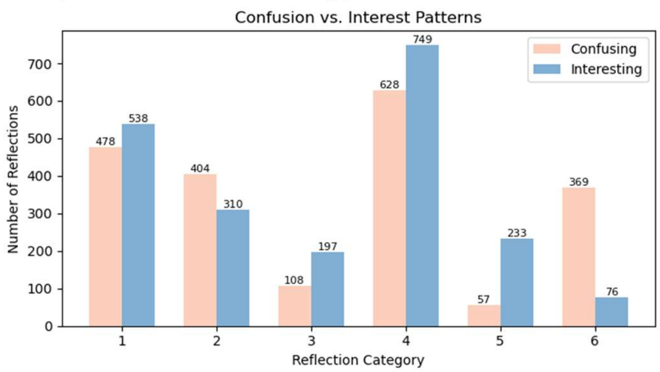

Figure 2. Confusion vs. Interest Patterns

Temporal Trends

Temporal analysis was conducted by combining both confusing and interesting reflections to examine overall reflection patterns across early, mid, and late stages. In the early phase, reflections were dominated by Technical Procedure and Instructional Context, indicating an initial focus on understanding course expectations and completing tasks. During the mid phase, there was an increase in Conceptual, Principle, and Miscellaneous reflections, suggesting deeper engagement with content alongside emerging uncertainty. In the late phase, Miscellaneous increased further, while Conceptual and Principle reflections decreased. This pattern may reflect accumulated unresolved confusion or general reflections as courses approached completion.

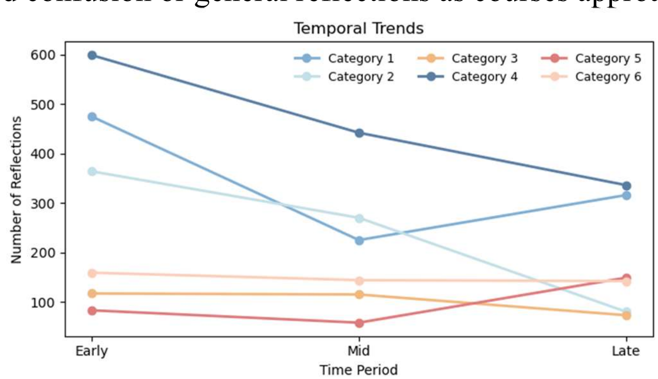

Figure 3. Temporal Trends

Cross-Course Comparisons

Both ENGR 101 and ENGR 102 were dominated by Technical Procedure and Instructional Context. However, ENGR 101 exhibited relatively higher levels of Metacognition, particularly in late-stage interesting reflections, indicating increased reflection on learning progress over time. In contrast, ENGR 102 displayed a more distributed pattern with higher proportions of Conceptual, Principle, and Miscellaneous reflections. A notable trend in ENGR 102 was the continued increase of Miscellaneous in confusing reflections across time.

Figure 4. Cross-Course Comparisons

Answers to Research Questions

RQ1: What topics or concepts are most frequently associated with confusion or interest in ENGR 101 and ENGR 102?

In ENGR 101, reflections most frequently focused on Technical Procedure and Instructional Context, indicating that students' confusion and interest were largely tied to executing tasks, using tools, and understanding course activities and expectations. Interest reflections in ENGR 101 showed relatively higher levels of Metacognition, suggesting that students increasingly reflected on their own learning processes as they became more familiar with the course structure. Confusion reflections, in contrast, included a substantial proportion of Miscellaneous entries, reflecting unresolved or broadly expressed uncertainty, especially around statistics-related course content. ENGR 102 exhibited a broader distribution of reflection categories compared to ENGR 101. While Technical Procedure and Instructional Context remained dominant, confusion reflections showed higher proportions of Conceptual, Principle, and Miscellaneous. This pattern reflects the increased cognitive demands of programming, where students often struggle to articulate specific misunderstandings [36].

RQ2: How do topics and concepts most frequently associated with confusion or interest overlap between ENGR 101 and ENGR 102?

Across both courses, Technical Procedure and Instructional Context consistently accounted for the largest share of reflections, showing challenges related to task execution and navigating course structures in first-year engineering. Meaningful differences emerged in how students articulated confusion and interest. ENGR 102 reflections showed greater emphasis on Conceptual and Principle, reflecting the increased demands of programming and computational reasoning. In contrast, ENGR 101 reflections more often centered on completing tasks and engaging with course activities. Together, these overlapping and divergent patterns indicate reflection content shifts as students progress through the first-year sequence, highlighting opportunities for instructors to more intentionally scaffold students' transition from statistical reasoning to computational problem-solving. Specifically, instructors may use reflection data to identify when students move from structured, formula-driven reasoning toward more abstract computational thinking, and to introduce targeted instructional supports that explicitly connect statistical concepts to programming constructs. Such scaffolding may help students recognize continuity across courses, reduce unarticulated confusion, and support more effective engagement with computational problem-solving tasks.

Limitations and Future Work

We acknowledge that there are limitations in this study that can be explored in future studies. One such limitation is that we conducted our coding at the reflection text level, which may cause issues with reflections that may contain multiple codes. An example of such is the following interesting reflection text:

"I found it interesting to see how to use the absolute and relative functions. It was helpful to see how we can directly apply these functions."

A coding discrepancy emerged for this text, as one coder classified this reflection as principle and the other classified this as technical procedure. Future studies could explore how LLMs can be leveraged to support researchers in identifying reflections that may encompass multiple codes and segment them accordingly. Future studies may also investigate how students may leverage LLMs to enhance their reflection specificity score and examine the potential implications for the learning process in reflection writing.

Another limitation of this study is that the data were drawn from two first-year engineering courses at one institution, which may limit the generalizability of the findings to other instructional contexts. In addition, the analysis focused on course-level reflection trends rather than tracking the same students across both courses. As a result, the findings reflect aggregate patterns rather than individual developmental trajectories as students' progress through the first-year engineering sequence. Future research can address these limitations by extending the analysis to multiple institutions, course sections, and instructors. Longitudinal studies that follow individual students across sequential courses could provide deeper insight into how reflection content evolves over time and how confusion and interest relate to learning development for students. In addition, future work will extend the analysis by examining reflection trends across different combinations of course, prompt type, and time period to provide a more detailed understanding of pattern variation. Finally, expanding the human-LLM coding approach to larger and more diverse datasets would further test the robustness and transferability of scalable reflection analysis in engineering education.

Acknowledgment
We thank Dr. Jason Morphew for providing valuable context about the course structure and supporting the data collection process.

References

[1] L. Sepp, M. Orand, J. Turns, L. Thomas, B. Sattler, and C. Atman, “On an Upward Trend: Reflection in Engineering Education,” in *2015 ASEE Annual Conference and Exposition Proceedings*, Seattle, Washington: ASEE Conferences, Jun. 2015, p. 26.1196.1-26.1196.13. doi: 10.18260/p.24533.

[2] L. Guo, “How should reflection be supported in higher education? — A meta-analysis of reflection interventions,” *Reflective Practice*, vol. 23, no. 1, pp. 118–146, Jan. 2022, doi: 10.1080/14623943.2021.1995856.

[3] C. Dignath, R. van Ewijk, F. Perels, and S. Fabriz, “Let Learners Monitor the Learning Content and Their Learning Behavior! A Meta-analysis on the Effectiveness of Tools to Foster Monitoring,” *Educ Psychol Rev*, vol. 35, no. 2, p. 62, May 2023, doi: 10.1007/s10648-023-09718-4.

[4] E. H. Eshuis, J. ter Vrugte, and T. de Jong, “Supporting reflection to improve learning from self-generated concept maps,” *Metacognition Learning*, vol. 17, no. 3, pp. 691–713, Dec. 2022, doi: 10.1007/s11409-022-09299-7.

[5] J. Kim, A. Satya Putra, E. T. R. Babar, M. Menekse, A. A. Butt, and S. Anwar, “Understanding Students Confusion and Interest in an Introductory Physics Course Through Qualitative Analysis of Self-Reflections,” 2025.

[6] A. A. Butt, S. Anwar, and M. Menekse, “WIP: Investigating the relationship between FYE students’ reflections and academic performance across gender,” presented at the 2022 First-Year Engineering Experience, Jul. 2022. Accessed: Sep. 21, 2025. [Online]. Available: https://peer.asee.org/wip-investigating-the-relationship-between-fye-students-reflections-and-academic-performance-across-gender

[7] H. A. Diefes-Dux and L. M. Cruz Castro, “Reflection types and students’ viewing of feedback in a first-year engineering course using standards-based grading,” *Journal of Engineering Education*, vol. 111, no. 2, pp. 283–307, 2022, doi: 10.1002/jee.20452.

[8] L. M. Cruz Castro, A. J. Magana, K. A. Douglas, and M. Boutin, “Analyzing Students’ Computational Thinking Practices in a First-Year Engineering Course,” *IEEE Access*, vol. 9, pp. 33041–33050, 2021, doi: 10.1109/ACCESS.2021.3061277.

[9] J. Brockman, T. Fuja, and S. Batill, “A Multidisciplinary Course Sequence For First Year Engineering Students,” presented at the 2002 Annual Conference, Jun. 2002, p. 7.69.1-7.69.20. Accessed: Jan. 15, 2026. [Online]. Available: https://peer.asee.org/a-multidisciplinary-course-sequence-for-first-year-engineering-students

[10] B. Frank, D. Strong, R. Sellens, and L. Clapham, “Progress with the Professional Spine: A Four-Year Engineering Design and Practice Sequence,” *Australasian Journal of Engineering Education*, vol. 19, no. 1, pp. 63–74, Jan. 2013, doi: 10.7158/22054952.2013.11464079.

[11] P. Burton, “The Challenge Of Teaching Large First Year Engineering Classes,” presented at the 2007 Annual Conference & Exposition, Jun. 2007, p. 12.1395.1-12.1395.7. Accessed: Jan. 15, 2026. [Online]. Available: https://peer.asee.org/the-challenge-of-teaching-large-first-year-engineering-classes

[12] C. Bovill, C. J. Bulley, and K. Morss, “Engaging and empowering first-year students through curriculum design: perspectives from the literature,” *Teaching in Higher Education*, vol. 16, no. 2, pp. 197–209, Apr. 2011, doi: 10.1080/13562517.2010.515024.

[13] D. A. Schon, *The Reflective Practitioner: How Professionals Think In Action*, 1st edition. New York: Basic Books, 1984.

[14] R. S. Adams, J. Turns, and C. J. Atman, "Educating effective engineering designers: the role of reflective practice," *Design Studies*, vol. 24, no. 3, pp. 275–294, May 2003, doi: 10.1016/S0142-694X(02)00056-X.

[15] J. Turns, K. Mejia, and C. Atman, "Reflection in Engineering Education: Advancing Conversations," in *2020 ASEE Virtual Annual Conference Content Access Proceedings*, Virtual On line: ASEE Conferences, Jun. 2020, p. 35131. doi: 10.18260/1-2--35131.

[16] G. van den Boom, F. Paas, and J. J. G. van Merriënboer, "Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes," *Learning and Instruction*, vol. 17, no. 5, pp. 532–548, Oct. 2007, doi: 10.1016/j.learninstruc.2007.09.003.

[17] M. Menekse *et al.*, "Enhancing Student Reflections with Natural Language Processing based Scaffolding: A Quasi-Experimental Study in a Large Lecture Course," *Computers and Education: Artificial Intelligence*, p. 100397, Mar. 2025, doi: 10.1016/j.caeai.2025.100397.

[18] H. J. Pijeira-Díaz, J. van de Pol, F. Channa, and A. de Bruin, "Scaffolding self-regulated learning from causal-relations texts: Diagramming and self-assessment to improve metacomprehension accuracy?," *Metacognition Learning*, vol. 18, no. 3, pp. 631–658, Dec. 2023, doi: 10.1007/s11409-023-09343-0.

[19] M. Fisher and D. E. Baird, "Making mLEarning Work: Utilizing Mobile Technology for Active Exploration, Collaboration, Assessment, and Reflection in Higher Education," *Journal of Educational Technology Systems*, vol. 35, no. 1, pp. 3–30, Sep. 2006, doi: 10.2190/4T10-RX04-113N-8858.

[20] T. Leinonen, A. Keune, M. Veermans, and T. Toikkanen, "Mobile apps for reflection in learning: A design research in K-12 education," *British Journal of Educational Technology*, vol. 47, no. 1, pp. 184–202, 2016, doi: 10.1111/bjet.12224.

[21] A. Knoth, A. Kiy, I. Müller, and M. Klein, "Competences in Context: Students' Expectations and Reflections as Guided by the Mobile Application Reflect.UP," *Tech Know Learn*, vol. 25, no. 4, pp. 707–731, Dec. 2020, doi: 10.1007/s10758-019-09407-8.

[22] S. Anwar, A. A. Butt, and M. Menekse, "Utilizing an NLP-supported mobile reflection application to explore academic engagement, application engagement, and performance in engineering and physics courses," *IJ STEM Ed*, vol. 12, no. 1, p. 41, Aug. 2025, doi: 10.1186/s40594-025-00551-5.

[23] X. Fan, W. Luo, M. Menekse, D. Litman, and J. Wang, "Scaling Reflection Prompts in Large Classrooms via Mobile Interfaces and Natural Language Processing," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, Limassol Cyprus: ACM, Mar. 2017, pp. 363–374. doi: 10.1145/3025171.3025204.

[24] R. Pekrun and E. J. Stephens, "Academic emotions," in *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors*, in APA handbooks in psychology®. , Washington, DC, US: American Psychological Association, 2012, pp. 3–31. doi: 10.1037/13274-001.

[25] K. Fayn, P. J. Silvia, E. Dejonckheere, S. Verdonck, and P. Kuppens, "Confused or curious? Openness/intellect predicts more positive interest-confusion relations," *Journal of Personality and Social Psychology*, vol. 117, no. 5, pp. 1016–1033, 2019, doi: 10.1037/pspp0000257.

[26] K. R. Muis, M. Chevrier, and C. A. Singh, "The Role of Epistemic Emotions in Personal Epistemology and Self-Regulated Learning," *Educational Psychologist*, vol. 53, no. 3, pp. 165–184, Jul. 2018, doi: 10.1080/00461520.2017.1421465.

[27] T. D. Ullmann, "Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches," *Int J Artif Intell Educ*, vol. 29, no. 2, pp. 217–257, May 2019, doi: 10.1007/s40593-019-00174-2.

[28] N. Than, L. Fan, T. Law, L. K. Nelson, and L. McCall, "Updating 'The Future of Coding': Qualitative Coding with Generative Large Language Models," *Sociological Methods & Research*, vol. 54, no. 3, pp. 849–888, Aug. 2025, doi: 10.1177/00491241251339188.

[29] C. K. Y. Chan and K. K. W. Lee, "Who Grades More Consistently? Exploring AI vs. Human Teachers in the Assessment of Reflective Writing," *AI Enhanced Learning*, vol. 1, no. 1, pp. 203–228, 2025, doi: https://doi.org/10.70725/298318rdotfx.

[30] X. Fan, W. Luo, M. Menekse, D. Litman, and J. Wang, "CourseMIRROR: Enhancing Large Classroom Instructor-Student Interactions via Mobile Interfaces and Natural Language Processing," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, Seoul Republic of Korea: ACM, Apr. 2015, pp. 1473–1478. doi: 10.1145/2702613.2732853.

[31] J. Saldaña, *The Coding Manual for Qualitative Researchers*, 3rd ed. Los Angeles ; London: SAGE Publications Ltd, 2015.

[32] M. Williams and T. Moser, "The Art of Coding and Thematic Exploration in Qualitative Research," *International Management Review*, 2019, Accessed: Jan. 15, 2026. [Online]. Available: https://www.semanticscholar.org/paper/The-Art-of-Coding-and-Thematic-Exploration-in-Williams-Moser/c0a0c26ac41cb8beb337834e6c1e2f35b91d071d

[33] M. T. H. Chi and K. A. VanLehn, "The Content of Physics Self-Explanations," *Journal of the Learning Sciences*, vol. 1, no. 1, pp. 69–105, Jan. 1991, doi: 10.1207/s15327809jls0101_4.

[34] J. Kim, M. Menekse, A. S. Putra, E. tur razia Babar, A. A. Butt, and S. Anwar, "Understanding Students' Confusion and Interest in an Introductory Physics Course Through Qualitative Analysis of Self-Reflections," presented at the 2025 ASEE Annual Conference & Exposition, Jun. 2025. doi: 10.18260/1-2--57741.

[35] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochem Med (Zagreb)*, vol. 22, no. 3, pp. 276–282, Oct. 2012.

[36] L. C. Kaczmarczyk, E. R. Petrick, J. P. East, and G. L. Herman, "Identifying student misconceptions of programming," in *Proceedings of the 41st ACM technical symposium on Computer science education*, in SIGCSE '10. New York, NY, USA: Association for Computing Machinery, Mar. 2010, pp. 107–111. doi: 10.1145/1734263.1734299.

Appendix A

Coding Scheme

Category Name	Description	Example Reflection(s)
Technical Procedure	Reflections about actions taken to execute the technical steps of the assignment/task (performing calculations, building something, running analysis, writing code) or perform system-level actions that enable the task (managing files, installing software, accessing platforms, navigating folders)	"I learned how to use the AVERAGE function in Excel to calculate mean values." "I had trouble saving my work to the right folder in Brightspace."
Conceptual Knowledge	Reflections about basic understanding of things, terms, or definitions.	"I now understand the difference between variance and standard deviation."
Principle-Related Knowledge	Reflections about how or why things work or the reasoning behind procedures and concepts	"We calculate standard deviation to see how spread out the data is."
Instructional Context	Reflections about course delivery, learning tasks (assignments, projects, exams, or activities), and how students practice, collaborate, or use learning resources.	"The project management assignment helped me see how engineers work in teams." "Working with classmates on the MATLAB lab made it easier to understand loops."
Metacognition	Reflections on one's own learning process, such as connecting to prior knowledge, recognizing mistakes, or monitoring progress.	"I learned this in high school, but today's explanation cleared up my misconceptions."
Miscellaneous	Reflections that are vague, general, or do not fit into other categories. This includes shallow expressions of confusion or interest.	"Today's lecture was confusing." "I thought the topic was interesting."

Appendix B

Instructions for the Customized Model

You are assisting in a qualitative analysis of student reflections collected from two first-year engineering courses (ENGR 101 and ENGR 102). Each reflection was written in response to one of two prompts: what the student found confusing or interesting during a lecture.

Your task is to assign one code to each reflection according to the finalized coding schema provided below. The schema follows an adaptation of Chi & VanLehn (1991) and includes categories such as Technical Procedure, Conceptual Knowledge, Principle-Related Knowledge, Instructional Context, Metacognition, and Miscellaneous.

Instructions:

- *Use only the codes and definitions from the provided codebook; do not add, rename, or merge categories.*
- *Review the provided human-coded examples (240 reflections) to understand how codes were applied.*
- *For each new reflection, determine the single best-fitting code that captures its main idea.*
- *If a reflection does not align with any of the other categories/simply states "nothing was confusing or interesting" without offering further details/does not relate to classroom activities or learning in first-year engineering, please categorize it as Miscellaneous.*
- *Provide a short rationale (1-2 sentences) explaining why you selected that code, referencing key phrases from the reflection when possible.*

Input:

- *Coding Schema.csv*
- *Coding Sample.csv*
- *Reflections to code (will provide specific information: which course is the reflection from, and what type of reflection is it (confusing or interesting), 20 reflections for one conversation)*

Output: Please produce a clean table with three columns:

- *Reflection (student reflection text)*
- *Code (the number of selected code)*
- *Rationale (brief reasoning citing reflection keywords)*

Appendix C

Example of Human Input and Corresponding LLM Output

Human Input:

The course: ENGR 101
The prompt type: Confusing
A set of 20 reflections to be coded:
1. Not necessarily confusing, but the EYM information was heavy and a bit overwhelming.
2. Some of the functions in excel need more clarifications.
…
19. The new formulas were pretty confusing. But got less confusing through practice.
20. Nothing was confusing. I need more explanation on what VAR and STDEV do. How does finding them help me gain more data in a chart?

LLM Output:

Below is the coded table for the 20 ENGR 101 Confusing reflections.

Reflection	*Code*	*Rationale*
Not necessarily confusing, but the EYM information was heavy and a bit overwhelming	*4*	*Refers to logistics and program requirements ("EYM information"), which fits the Instructional Context.*
Some of the functions in excel need more clarifications.	*1*	*"Functions in Excel" refers to performing procedures in software → Technical Procedure.*
…		
The new formulas were pretty confusing. But got less confusing through practice.	*1*	*"New formulas" in Excel implies procedural difficulty → Technical Procedure.*
Nothing was confusing. I need more explanation on what VAR and STDEV do. How does finding them help me gain more data in a chart?	*2*	*Asks about conceptual purpose of variance & stdev ("what … do", "how does … help") → Conceptual Knowledge.*