2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51697

Developing critical thinking skills in engineering students through a primary literature exercise

Dr. Joshua Kays, Boston University

Developing critical thinking skills in engineering students through a primary literature exercise

Abstract

Critical thinking (CT) skills are essential to undergraduate engineering education, yet despite the association of an engineering degree with CT ability, studies indicate that traditional engineering curriculum fail to create gains in students' CT aptitude. This study introduces an innovative structured "deep dive" literature analysis exercise aimed at enhancing CT skills in engineering students. Conducted over a semester with a graduate-level biomedical engineering class, the exercise focused on interpreting, analyzing, and questioning primary literature. Using a mixed-methods approach, we assessed students' perceived CT abilities across six dimensions using a shortened form of the Critical Thinking Self-Assessment Scale (CTSAS) and open-ended responses. Results indicate a significant improvement in students' perceived CT skills, with 87-93% of participants reporting enhanced abilities across all dimensions of CT. These exercises appeared equally helpful across gender, indicating the deep dives do not induce inequality. Additionally, qualitative analysis of open-ended responses revealed positive emotional engagement and highlighted specific CT skills gained, including improved evaluation, inference, and explanation. This exercise advances engineering pedagogy by offering a readily adaptable tool to improve their students' CT skills while covering course-appropriate content through primary literature.

Introduction

Critical thinking (CT) skills have long been a focus of undergraduate education, and in the US critical thinking was explicitly part of ABET accreditation criteria since 2000. While multiple models and definitions of CT exist, an expert consensus (identified by the Delphi method) characterized six "dimensions" of CT: interpretation, analysis, evaluation, inference, explanation, and self-regulation [1] as seen in the appendix (Appendix Table 1). This framework and its dimensions has been used across multiple fields and is the basis for multiple CT assessments, including the California Critical Thinking Disposition Inventory (CCTDI)[2] and the Critical Thinking Self-Assessment Scale (CTSAS)[3], [4].

Engineering students and educators often believe that because engineering classes teach problem solving, critical thinking skills are inherently taught and covered (for a student example, see [5]); however, the problems offered in engineering curriculum are often highly bounded, lack ambiguity, and can be solved via a simple "plug and chug" method[6]. This "algorithmic progression through problem-solving stages"[6] does not necessarily build essential critical thinking skills. Furthermore while critical thinking skills are highly sought after across multiple industries[7], [8], employers increasingly believe that engineering students are not sufficiently equipped to think critically and independently[9], [10]. Indeed, in one study by Sola *et al*, 4th year engineering students showed no increase in critical thinking skills compared to 1st year, indicating a gap in their training[11]. Thus, there is a strong need for techniques and activities for engineering students that improve critical thinking skills beyond solving problem sets. This is especially relevant for biomedical engineering (BME), since it is highly interdisciplinary and engages with highly complex and context-dependent biological systems[12].

Reviews of the literature have shown that most exercises to improve critical thinking are either projects or "real-world" problems and case studies[13]. While these are helpful, they can be time consuming, work best with design classes, and may rely on experiences that professors may not have in their academic careers. Primary literature offers an accessible and low barrier to entry option for educators. As most engineering professors are academics, they are familiar with and have expertise in reading and analyzing primary literature. Additionally, academic literature has *already* been used to improve critical thinking in non-engineering domains[14], [15], [16], [17], yet surprisingly, this has not been extended to an engineering context.

In this study, I developed a "deep dive" literature analysis exercise, conducted this exercise 4 times, and subsequently measured student's assessment of their critical thinking skills using shortened form of the Critical Thinking Self-Assessment Scale (CTSAS)[4], [18], based on the along with open-ended questions. This exercise combines written analysis, verbal discussion, and multiple types of formal and informal feedback, giving students many opportunities to engage with the material. This study aims to (a) describe the design of this novel literature analysis exercise, and (b) assess the impact of this exercise on student's perceived critical thinking abilities across six dimensions of critical thinking using a mixed-methods approach.

Methods

Participants and course content: This study was conducted in a graduate level biomedical engineering elective class (n=21 students) at a large private university in the USA northeast region. This 700-level course titled "Nanomedicine" is designed for Ph.D. students and contained three major stages: fundamentals of nanomedicine, bio-nano interactions across cellular, tissue, and system levels, and clinical applications. The class contained 8 Ph.D. students, 10 masters students, and 3 undergraduate students, with an approximately equal split of students by sex (57% male, 43% female). Demographically the majority of students were East Asian or Indian (12 students), followed by Caucasian (6), Latin American (2) and African American (1). Students opted into all surveys and data collection, and data was collected anonymously. Gender was collected by the survey, while all other demographic information was obtained through university data and was not linked to responses. This study was reviewed beforehand by the university's Institutional Review Board (Application #7489X) and deemed exempt.

Activity design: The deep dive exercises were conducted as follows. In the 3rd class (out of 26) of the semester, students were trained to think critically and analyze a journal article (Figure 1A) in the area of nanomedicine. Specifically, they were taught a framework of interpreting, analyzing, and questioning three aspects of the article – a) the problem they claim to solve, b) the method they go about to demonstrate their solution, and c) the validity and effectiveness of their solution.

This framework was designed to include at least five of the six dimensions of CT; however, students were <u>never explicitly taught the six dimensions of CT</u>. For example, students were taught to check both axes of a plot, to look at the trend of the data, ask what "story" the data is telling, and write down their answers to that question. These are all demonstrations of "interpretation", the first of the six dimensions. In another example, under questioning the "problem", students were shown claims in the introduction of a journal article that set up the problem, and then were asked a series of questions about the accuracy and saliency of these citations to the original claim. This led students to see that sometimes citations would be inaccurate or inappropriate, demonstrating to the students the skills of both "evaluation" and "inference" (3rd and 4th dimension of CT).

After this initial training class, students were given a new journal article to conduct their own "deep dive" over the course of two weeks (Figure 1B, top) These articles were sourced from top nanomedicine journals such as *ACS Nano* and *Nature Nanotechnology*. Students were encouraged to discuss with each other but instructed to write up their own analysis and critique of the article (in 600-1200 words), using the same framework of problem, method, and solution. In the assignment, students were explicitly instructed to avoid summarizing the journal article, but instead to use claims, evidence, and logic to argue against or for the article's claims. These essays were read by the instructor (or peer) at a subsequent date; feedback was given, and assignments were graded based on a rubric by either the instructor, a peer, or self-grading.

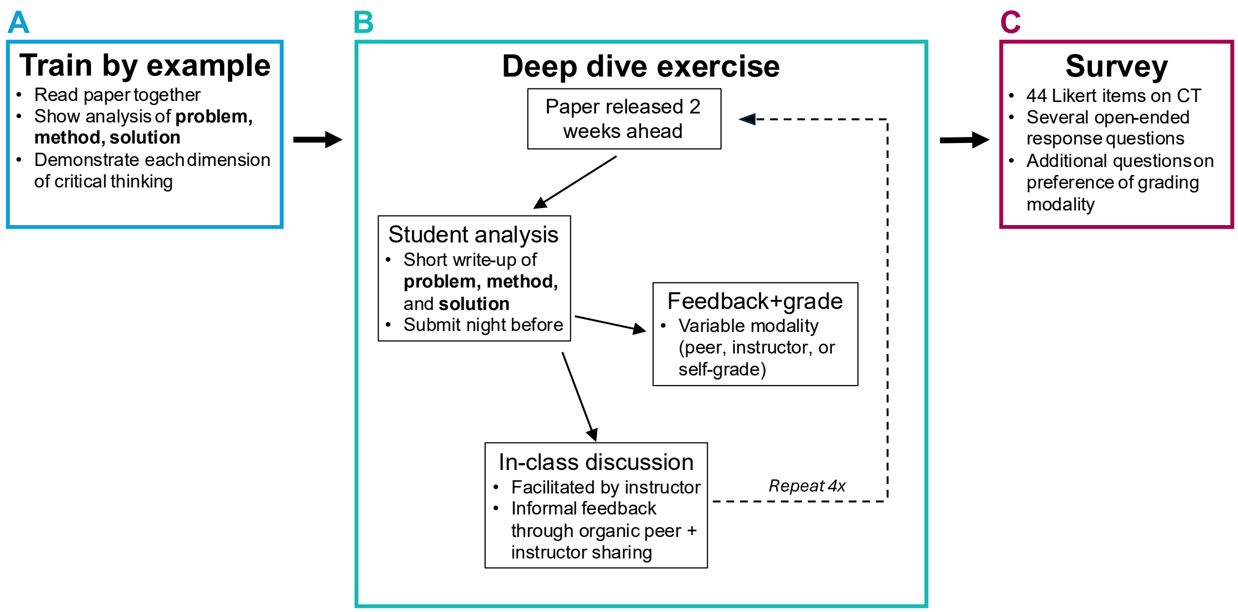

Figure 1: Diagram of the deep dive exercises

The day after the assignment due date a discussion was held during the 105min class period (Figure 1B, bottom). During this discussion students shared their overall thoughts or impressions of the article, as well as their main reactions to each figure and to the overall problem, method, and solution of the article. To encourage participation, think-pair-share was often employed. Additionally, real-time surveys using QR codes to google forms were employed to gather student responses to specific questions. This discussion was intentionally structured to enable students to hear multiple perspectives on the claims of the article, especially given the diversity of educational backgrounds in the room (undergraduates through mid-stage Ph.D. students). This deep dive exercise was repeated 4 times over the course of the semester, after which the survey was given (Figure 1C).

Survey design and analysis: In this study, I utilized a shortened form of the CTSAS adapted from Payan-Carreira *et al*[18], whose article contains the CTSAS questions. This 5-point Likert-scale self-assessment included 44 questions spread across the six dimensions of CT described above, as well as two open-ended questions to further understand their experience. Additional Likert scale questions were asked regarding grading and overall impact of the exercises on their thinking. The survey was administered after receiving student consent through Qualtrics software, with question randomization and no identifying information collected. Likert items were analyzed using descriptive statistics, where two-way T-tests with unequal variances were used to compare differences by gender and ANOVA (Analysis of Variance, $\alpha = 0.05$) was performed for assessing differences between critical thinking dimensions. Sentiment and content analysis of the open-ended responses were conducted with AI. AI based on Large Language Models (e.g. ChatGPT, Copilot) is best used for this task, as they can accurately code responses and parse intent and tone with less bias than human authors[19], [20], [21]. Sentiment scores were determined based on the overall tone and content of each feedback, on a -9-point scale (from -4 to 4, where 0 is fully neutral, 1 = slightly positive, 2 = moderately positive, 3 = very positive, and 4 = highly positive). Inductive content analysis was performed first to identify unique but coarse-grained categories, such that the number of categories was less than the number of respondents. Then the coverage of each response was calculated – where coverage is defined as text of each response that was coded / total text of each response – and checked to ensure that 50% or more each student's response was coded by at least one or more categories. After any corrections, a thematic network analysis was performed on the categories to identify broad global themes. Data was first analyzed with k-means clustering in MATLAB to determine how many unique themes should be used: this was checked by qualitatively generating key themes. Category codes were then assigned to each theme. The frequency and co-occurrence (as Jaccard index) of themes in student responses was determined using MATLAB software. The Jaccard index was used as it is a robust measure of similarity[22], [23].

Results and Discussion

Quantitative self-assessment: Student responses to the Likert survey demonstrated that they

believe these exercises improved their CT skills (Figure 2). Specifically, between 87% - 93% of students believed the deep dive exercises improved their CT across all dimensions, with 72% or more strongly agreeing for the first five dimensions (Figure 2A). While visually students responded slightly less positive for the last CT dimension of "self-regulation", results from the ANOVA (Appendix Table 2) demonstrate that there was no significant difference between the six dimensions (p-value = 0.247). Pairwise T-tests demonstrated no difference between respondents by gender (Figure 2B), indicating that these exercises did not induce biased outcomes between men and women.

It is worth noting that the higher variability and slightly lower score for the men was due to one outlier respondent who may have misread the scales, as his open-ended responses were highly positive but Likert item responses all highly negative – in fact, his answers account for all the "strongly disagree" in 2A. When that outlying respondent is removed, the difference between means reduces to <2% across all dimensions and variability drops significantly. To confirm that the Likert scales for each dimension of CT were internally consistent, Cronbach's alpha was calculated for each. Four of the dimensions yielded Cronbach's alpha of >0.95 and all were ≥0.88, indicating an excellent internal consistency in each dimension.

Students' critical thinking – open-ended responses: To further examine the students' perception of their critical thinking skills, they were asked to "Please share any reflections on the deep dive exercises and their impact on your critical thinking and analysis skills" in an unbounded open-ended format; 18 out of 21 students responded. As a whole, students' open-ended responses matched their favorable ratings in the Likert scales (except for 1 student, as previously mentioned). The sentiment analysis, scored from -4 to 4, yielded no neutral or negative responses: 15 of the 18 were very positive or higher, with the average score of 3.33 (Figure 3). A few illustrative responses are included below:

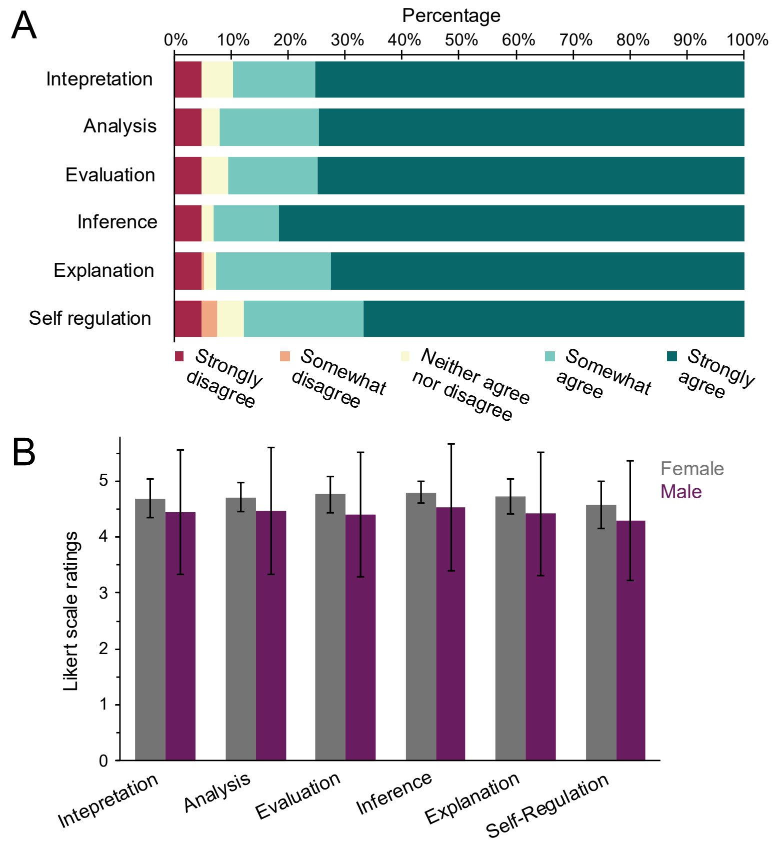

Figure 2: Analysis of Likert scale survey responses.
A) Percentage of responses across each dimension of CT.
B) Mean response rating by gender. There was no significant difference between gender for any CT dimension. Error bars represent one S.D.

- Sentiment score 4: "*The deep dives have helped me tremendously. I believe they are especially helpful when it comes to thinking critically and analyzing a new piece of data in a logical framework, which is universally applicable. I think is invaluable as a skill to have. I have never self-evaluated my problem-solving approach, but now that I have my approach has completely changed. And, I believe its for the better as I find myself being more reticent to make claims or readily listen to others claims without evidence to support them.*"

- Sentiment score 3: "*Certainly improved my critical thinking skills. Never knew I was 'allowed' to question papers and search for flaws in their reasonings. Used to believe in the methodologies of everything written in papers. It was fun showcasing my thoughts on the 'problems' and 'solutions', how some problems are actually not worth solving, or the results didn't solve the problem.*

- Sentiment score 4: *"I came into the class thinking I knew how to read papers and realized that I was never really critically engaging with them. If a paper claimed X, to me that was fact. Having participated in these deep dive exercises, I can confidently say that my skills have improved and have helped me read papers outside of nano. Aside from reading papers critically, I also feel that I know what questions to ask when I am learning something and to challenge ideas that are new to me. In a way I feel I have grown into a bit of a skeptic, but in a positive way where I want to truly understand how someone came to a conclusion or what lead them to that claim. It has improved my critical thinking overall, and has been an invaluable experience."*

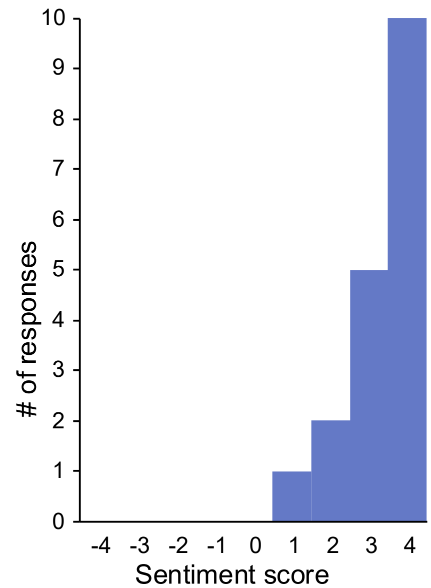

Figure 3: Histogram of the sentiment analysis scores

As seen in these responses, students seemed to enjoy these deep dive exercises and believed them to be "invaluable", while also indicating how they improved in their critical thinking skills. This survey data matches students' verbal comments to the instructor as well as comments collected by the University's standard course evaluation, where students shared that this activity increased their understanding of, engagement with, and even enjoyment of biomedical literature.

To better measure the student's responses, a semi-quantitative content analysis was performed. First, eight unique but coarse-grained codes were developed inductively from the responses. The responses were then coded (n=42 codes applied to 18 responses) to ensure sufficient coverage: indeed, coverage ranged between 58% to 99% with an average of 83%. These codes were then clustered into 4 overarching themes – "Feedback", "Practical Relevance", "Emotional Engagement ", and "CT Skill Enhancement" (Figure 4A).

The first two themes – "Feedback" and "Practical Enhancement" – represented the fewest responses (n=3 and n=2, respectively) and also the least co-occurrence with other themes (all values less than 0.15, see figure 4B. Feedback relates to students sharing feedback on (or suggestions for) the exercise, while Practical Enhancement contains ways the students felt the exercises directly prepared them for something practical – namely, job interviews and the PhD qualifying exam. The most substantive feedback as it pertains to CT was a comment that "the way we discussed papers tended to reward negative critique over appreciation for quality work."

Though the student follows up with "though this may not be a bad thing", their feedback hints at an area of improvement in both facilitation the discussion and possibly in the teaching of the deep dive exercises in general. The two remaining comments were unrelated to CT and to each other: one student desired a different selection of papers, and the other wished that mutual grading was not used for one of the deep dive assignments. The Feedback theme's low co-occurrence indicates that this theme was not related to other themes and most likely represents isolated ideas from students, in contrast to a clear pattern of feedback. Finally, the thematic clustering matched well with the sentiment analysis: two of the three lowest scores in the sentiment analysis were contained in the Feedback theme.

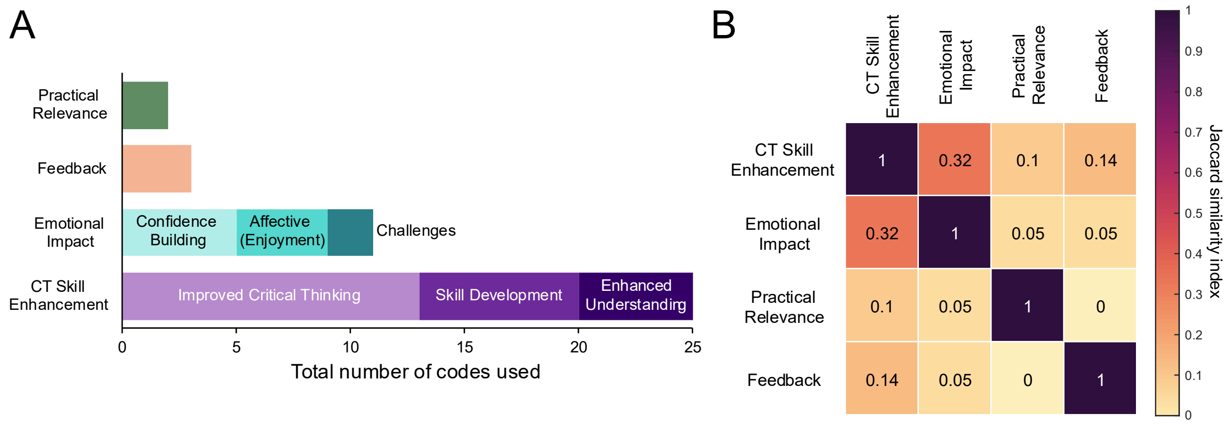

Figure 4: Thematic Analysis of open-ended responses. A) Frequency of themes (total) and codes (constituent parts) from content analysis and post-analysis thematic clustering. Both "Practical Relevance" and "Feedback" themes contained only one code, and thus the themes are named by the code used. In contrast, the bottom two themes contain three codes each. B) Co-occurrence matrix (as a heatmap) of the four themes in student responses. The values and color both represent the calculated Jaccard index.

The second largest theme was that of "Emotional Impact", which included codes for Affective Responses (Enjoyment), Confidence building, and Challenges. The positive comments in response to a neutral sharing prompt speak to the students' enjoyment of each aspect of the deep dives, be it paper reading ("I really enjoyed reading the deep dives"), writing their analysis ("It was fun showcasing my thoughts on the 'problems' and 'solutions' [of each paper]"), or discussion ("I enjoyed…hearing the opinions of my classmates"). Even in the three responses coded "Challenges", two of them spoke of the exercise in a positive light (sentiment score =3): one mentioned the *initial* challenge of the deep dives but discussed their subsequent growth in them, while the other desired to spend more time on the deep dives but was unable due to other work. Multiple students (n=5) shared that the deep dives increased their confidence in analyzing or thinking deeply. The positive affective response these exercises generated in students is significant: while it is increasingly well known that emotional engagement is a highly important aspect of student learning and development[24], [25], [26], there has been little research into measuring affective engagement or designing strategies to enhance it[27]. The deep dives provide an opportunity for that positive emotional engagement. This is perhaps due to the exercises' emphasis on students' own thoughts and justifications for their thoughts, which may give a sense of "ownership" to the students. In keeping with this research, the relatively high co-occurrence (Figure 4B) between this theme and the "CT Skill Enhancement" theme indicates that students who expressed positive emotional impact also expressed increased CT skill enhancement, suggesting a correlation between the two. In short, this theme confirms the strong positive sentiment analysis and shows that the students found these exercises enjoyable and worthwhile, even when challenging.

The final and largest theme by both number of codes (n=25) and number of students (n=17 out of 18) was the "CT Skill Enhancement", which encompassed codes for Improved Critical Thinking, Enhanced Understanding, and Skill Development. While these three codes may appear similar, they reflect the varied ways students self-identified their development: some focused on specific skills (ex: the ability to generate a conclusion), others on critical analysis and interpretation, and still others referred to a general growth in their perspective and understanding of a given topic. As discussed previously, this theme had significant co-occurrence with the Emotional Impact theme (Jaccard index = 0.32, see Figure 4B). This theme compliments the Likert responses (Figure 2) in two ways. Firstly, that 17/18 students expressed how this exercise positively impacted their CT abilities seems to add to the validity of the quantitative Likert data. Secondly and surprisingly, all six dimensions of CT seem to appear in some form in the open-ended data, despite students never being explicitly taught these dimensions. Some examples are given below in Table 1 below

Table 1: CT dimensions demonstrated in students' responses

CT Dimension	**Representative quote(s) that demonstrates a given CT dimension**	**Reasoning**
Interpretation	"I do think the deep dives helped me better understand the papers not only on their own but in the context of nanomedicine [literature]"	Demonstrates comprehension of the meaning and significance of papers in the context of nanomedicine.
Analysis	"I learned how to critically assess literature, as well as how to constructively discuss pros and cons about research." "I was initially very scared about reading scientific literature, but being able to divide them into modes of critical reasoning (problem, methods, solution) really helped my analysis. "	The first quote reflects the ability to identify concepts in the paper and their relationships, and to articulate the strengths and weaknesses of the content. The second quote demonstrates categorizing the paper's contents into subdomains (Facione 2010)
Evaluation	"The deep dives really helped me to read papers better, and actually question the claims and data presented in a paper, rather than just believe whatever they claim." "The deep dives allow me to establish the skills to challenge the integrity of scientific research."	Both quotes demonstrate the skill of assessing the credibility of the claims and data.
Inference	"I couldn't generate a conclusion after reading a paper before, but now I can. And I could interpret the figures better." "The deep dive exercises greatly improved my ability to question the overall claims and 'evidence' that authors present in papers."	The first quote demonstrates the ability to draw conclusions and infer meaning from the paper and its figures. The second quote shows the ability to determine the author's main claims and evidence, as well as aspects of Evaluation (see above).

Explanation	"The entire exercise of the deep dives were really impactful for me as they …helped me articulate my critiques and thoughts" "The deep dive exercises also improved my ability to explain my thinking process in writing…"	Both respondents describe improved ability to explain their thinking in written form, demonstrating the presentation of arguments and justifying reasoning, key aspects of explanation.
Self-Regulation:	"I have never self-evaluated my problem-solving approach, but now that I have my approach has completely changed."	The respondent is analyzing their problem-solving approach and reflecting on the change (self-regulation),which is evidence of self-correction and self-regulation.

That all six dimensions of CT were mentioned or demonstrated in a singular open-ended response question points to the potential efficacy of the deep dive exercises. An exciting future work would be to judge students' papers quantitatively or qualitatively over time to explicitly demonstrate their improvement in CT – for example, by training a LLM on the dimensions of CT through the CCTDI, then feeding it the papers to judge and score.

Impact beyond the classroom In addition to the Likert questions about CT, students were also asked via Likert-scale about how the deep dive exercises impacted two other domains: a) their ability to communicate their opinion and reasoning in writing, and b) how they approach 'everyday' information (conversations, news, social media, essays, etc). In both cases, students responded overwhelmingly in the affirmative (Figure 5). All students believed their "ability to communicate [their] opinion and reasoning" improved because of the deep dives, with over 70% believing they were "much better". Written communication is a highly sought-after skill, so the students' perception of growth in this area adds to the classroom value of this exercise. Excitedly, 96 students also believed this exercise positively changed how they interacted with 'everyday' information, such as news, conversations, social media, essays, and so forth. (Figure 5B).

A few selected responses below elaborate on the Likert data:

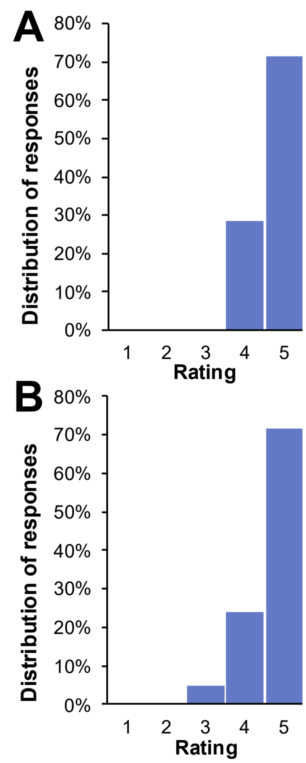

Figure 5: Distribution of students' responses to two questions. A) Responses to whether the deep dives improved written communication ability, scale of 1-5 reflecting much worse or much better. B) Agreement with the following: "The deep dives have changed how I approach everyday information (news, conversations, social media, essays, etc) for the better"

- *I always had heard terms like "fake news" when describing something that someone thought was untrue. However, I did not have the best ability to analyze what I was actually reading and think about how I understand it. I now also see how half-truths, assumptions, and omissions can radically change my feelings on a particular topic.*

- *These deep dive assignments have taught me to be better at critically questioning and evaluating information of all kinds.*

- *I think the deep dive exercises gave me the fundamentals to critically think about anything. Hearing others ask questions about papers, ideas, content during my time in grad school thus far, I've always wondered to myself "how did they come up with that question?" and "how do I know what questions to ask?" The deep dive exercises not only challenged me to think critically, but also provided me with the guidance and fundamentals to know what to question, how to question it, and how to draw connections. It also gave me the confidence to draw conclusions of my own and to not blindly trust claims just because an author of a paper said so. This has translated to claims of all kinds, from news to casual conversations I have with people.*

Limitations: While the results of this work are exciting, this study has several limitations that should be acknowledged. The first is the small class size: only 21 students in total. This limited sample size may affect the generalizability of the results to larger student populations. Future studies with larger cohorts are necessary to validate these findings. Additionally, the class was composed of primarily graduate-level students, which raises questions on its applicability to undergraduate students. However, I note that an earlier iteration of the "deep dive" approach was used on an undergraduate only class composed of mostly 3rd year students, which was well received and well-reviewed in standard University course evaluations, suggesting that this approach could be translated to other classes.

Another limitation is that the exercises were applied within the context of one specific biomedical engineering course (e.g. nanomedicine). While the framework of problem—method—solution is versatile and could be applied across multiple electives and disciplines, future studies should explore its effectiveness in other areas of engineering and science to confirm its broader applicability. Furthermore, this study was conducted over a single semester, which may not be sufficient to capture the long-term impact of the exercises on students' critical thinking development. Ideally, critical thinking should be cultivated through a variety of techniques over a more extended period, as argued in Ahern *et al*[13]. However, implementing intentional, long-term, and widespread changes across the undergraduate educational experience necessitates the buy-in of high-level academic administrators (such as deans and provosts) and would require significant institutional support. These are hard constraints that are not easily met.

This study primarily examined self-assessment of critical thinking, relying only on the framework of Facione. While this framework is well-established, other critical thinking frameworks – such as those of Paul and Elder[28] – and their associated assessments were not utilized. Future research should consider using multiple frameworks to provide a more comprehensive evaluation of students' critical thinking abilities. The greatest limitation – in the opinion of this author – is the reliance on self-assessment, increase of quantifying critical thinking with some validated tool (such as the Watson Glaser Critical Thinking Exam). However, these tools are typically expensive (>$30 per exam per person) and are less accessible to researchers. They also have their own limitations, and there is a need for novel accurate and validated assessments for researchers to use[29]. I will note that student's grades on their deep dives, which were judged on the depth of their analysis, improved from an average of 74 on the first deep dive to an average of 95 by the fourth deep dive.

Conclusion

The "deep dive" literature analysis exercises implemented in this study have demonstrated significant potential in enhancing critical thinking (CT) skills among engineering students. By engaging students in a structured framework that involves interpreting, analyzing, and questioning primary literature, I have observed a marked improvement in their ability to critically evaluate scientific claims. The survey results, both quantitative and qualitative, strongly indicate that students not only recognized an increase in their CT abilities but also believed themselves to be better prepared to think critically in in real-world settings. The high levels of student engagement and positive emotional responses suggest that such exercises can effectively foster a deeper understanding and appreciation of critical thinking. Additionally, the lack of significant gender-based differences in survey responses highlights the inclusive nature of this pedagogical approach.

Though perhaps appearing similar to traditional journal clubs from a cursory glance, the "deep dive" literature analysis exercises differ significantly. The framework used by the deep dives – examine the problem it claims to solve, the method it goes about solving it, and the solution it proposes – systematically guides students through interpreting, analyzing, and questioning of primary literature, yet this framework is generalizable to whitepapers, essays, and other forms of persuasive argumentation. The exercises also incorporate multi-modal engagement - including written analysis, verbal discussion, and iterative feedback - fostering deeper affective and cognitive engagement. Additionally, these exercises emphasize constructing well-founded arguments and provide measurable, evidence-based assessments of their effectiveness, demonstrating broader applicability beyond academic contexts.

While the small sample size and the single-institution context limit the generalizability of these findings, the promising outcomes warrant further exploration and application across different contexts. Future studies could extend this approach to undergraduate courses and other engineering disciplines, or extend across multiple classes. It would be especially interesting to design a way to score student's critical thinking in their written analysis and track their improvement over time in a quantitative way.

Ultimately, integrating primary literature analysis into engineering curricula offers a practical and scalable method to cultivate essential critical thinking skills, thereby preparing and empowering students for the complex problem-solving demands of their future careers.

Bibliography

[1] P. A. Facione, "Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations," 1990. Accessed: Jun. 10, 2024. [Online]. Available: https://eric.ed.gov/?id=ED315423

[2] N. C. Facione, P. A. Facione, and C. A. Sanchez, "Critical Thinking Disposition as a Measure of Competent Clinical Judgment: The Development of the California Critical Thinking Disposition Inventory," *Journal of Nursing Education*, vol. 33, no. 8, pp. 345–350, Oct. 1994, doi: 10.3928/0148-4834-19941001-05.

[3] G. G. Nair and L. L. Stamler, "A Conceptual Framework for Developing a Critical Thinking Self-Assessment Scale," *Journal of Nursing Education*, vol. 52, no. 3, pp. 131–138, Mar. 2013, doi: 10.3928/01484834-20120215-01.

[4] G. G. Nair, L.-A. M. Hellsten, and L. L. Stamler, "Accumulation of Content Validation Evidence for the Critical Thinking Self-Assessment Scale," *J Nurs Meas*, vol. 25, no. 1, pp. 156–170, Apr. 2017, doi: 10.1891/1061-3749.25.1.156.

[5] M. Tran, "Critical Thinking for Engineers | Electrical and Computer Engineering Design Handbook." Accessed: Jun. 10, 2024. [Online]. Available: https://sites.tufts.edu/eeseniordesignhandbook/2013/an-engineers-path-to-critical-thinking/

[6] T. Kenyon, "Critical thinking for engineers and engineering critical thinking," in *2016 2nd International Conference of the Portuguese Society for Engineering Education (CISPEE)*, Oct. 2016, pp. 1–4. doi: 10.1109/CISPEE.2016.7777736.

[7] M. Hirudayaraj, R. Baker, F. Baker, and M. Eastman, "Soft Skills for Entry-Level Engineers: What Employers Want," *Education Sciences*, vol. 11, no. 10, Art. no. 10, Oct. 2021, doi: 10.3390/educsci11100641.

[8] L. Claris and D. Riley, "Situation critical: critical theory and critical thinking in engineering education," *Engineering Studies*, vol. 4, no. 2, pp. 101–120, Aug. 2012, doi: 10.1080/19378629.2011.649920.

[9] D. Dumitru *et al.*, *A European Collection of the Critical Thinking Skills and Dispositions Needed in Different Professional Fields for the 21st Century*. 2018.

[10] I. Gashurov and A. Matsuuchi, "CUNY's Critical Thinking Skills Initiative: Redesigning workforce education through information literacy learning," *crln*, vol. 74, no. 2, pp. 70–73, Feb. 2013, doi: 10.5860/crln.74.2.8897.

[11] E. Sola, R. Hoekstra, S. Fiore, and P. McCauley, "An Investigation of the State of Creativity and Critical Thinking in Engineering Undergraduates," *Creative Education*, vol. 8, no. 9, Art. no. 9, Jul. 2017, doi: 10.4236/ce.2017.89105.

[12] K. A. Janes *et al.*, "An engineering design approach to systems biology," *Integr Biol (Camb)*, vol. 9, no. 7, pp. 574–583, Jul. 2017, doi: 10.1039/c7ib00014f.

[13] A. Ahern, C. Dominguez, C. McNally, J. J. O'Sullivan, and D. Pedrosa, "A literature review of critical thinking in engineering education," *Studies in Higher Education*, vol. 44, no. 5, pp. 816–828, May 2019, doi: 10.1080/03075079.2019.1586325.

[14] T. Li and Y. Liu, "Critical Thinking through Literature Review in Organic Chemistry Laboratory Course," *J. Chem. Educ.*, vol. 99, no. 7, pp. 2572–2578, Jul. 2022, doi: 10.1021/acs.jchemed.2c00063.

[15] S. G. Hoskins, "Using a Paradigm Shift to Teach Neurobiology and the Nature of Science—a C.R.E.A.T.E.-based Approach," *J Undergrad Neurosci Educ*, vol. 6, no. 2, pp. A40–A52, Jun. 2008.

[16] S. G. Hoskins, D. Lopatto, and L. M. Stevens, “The C.R.E.A.T.E. Approach to Primary Literature Shifts Undergraduates’ Self-Assessed Ability to Read and Analyze Journal Articles, Attitudes about Science, and Epistemological Beliefs,” *CBE Life Sci Educ*, vol. 10, no. 4, pp. 368–378, 2011, doi: 10.1187/cbe.11-03-0027.
[17] B. K. Sato, P. Kadandale, W. He, P. M. N. Murata, Y. Latif, and M. Warschauer, “Practice Makes Pretty Good: Assessment of Primary Literature Reading Abilities across Multiple Large-Enrollment Biology Laboratory Courses,” *LSE*, vol. 13, no. 4, pp. 677–686, Dec. 2014, doi: 10.1187/cbe.14-02-0025.
[18] R. Payan-Carreira, A. Sacau-Fontenla, H. Rebelo, L. Sebastião, and D. Pnevmatikos, “Development and Validation of a Critical Thinking Assessment-Scale Short Form,” *Education Sciences*, vol. 12, no. 12, Art. no. 12, Dec. 2022, doi: 10.3390/educsci12120938.
[19] F. Gilardi, M. Alizadeh, and M. Kubli, “ChatGPT outperforms crowd workers for text-annotation tasks,” *Proc Natl Acad Sci U S A*, vol. 120, no. 30, p. e2305016120, Jun. 2023, doi: 10.1073/pnas.2305016120.
[20] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, and D. Yang, “Is ChatGPT a General-Purpose Natural Language Processing Task Solver?,” in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, and K. Bali, Eds., Singapore: Association for Computational Linguistics, Dec. 2023, pp. 1339–1384. doi: 10.18653/v1/2023.emnlp-main.85.
[21] P. Törnberg, “How to use LLMs for Text Analysis,” Jul. 24, 2023, *arXiv*: arXiv:2307.13106. Accessed: Jun. 26, 2024. [Online]. Available: http://arxiv.org/abs/2307.13106
[22] K. Sdravopoulou, J. M. Muñoz González, and M. D. Hidalgo-Ariza, “Educating Adults with a Location-Based Augmented Reality Game: A Content Analysis Approach,” *Mathematics*, vol. 9, no. 17, p. 2071, Jan. 2021, doi: 10.3390/math9172071.
[23] A. Park, M. Conway, and A. T. Chen, “Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach,” *Computers in Human Behavior*, vol. 78, pp. 98–112, Jan. 2018, doi: 10.1016/j.chb.2017.09.001.
[24] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, “School Engagement: Potential of the Concept, State of the Evidence,” *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, 2004.
[25] R. Pekrun and L. Linnenbrink-Garcia, “Academic Emotions and Student Engagement,” in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds., Boston, MA: Springer US, 2012, pp. 259–282. doi: 10.1007/978-1-4614-2018-7_12.
[26] Z. Y. Wong and G. A. D. Liem, “Student Engagement: Current State of the Construct, Conceptual Refinement, and Future Research Directions,” *Educ Psychol Rev*, vol. 34, no. 1, pp. 107–138, Mar. 2022, doi: 10.1007/s10648-021-09628-3.
[27] R. Maguire, A. Egan, P. Hyland, and P. Maguire, “Engaging students emotionally: the role of emotional intelligence in predicting cognitive and affective engagement in higher education,” *Higher Education Research & Development*, vol. 36, no. 2, pp. 343–357, Feb. 2017, doi: 10.1080/07294360.2016.1185396.
[28] P. Caratozzolo, A. Alvarez-Delgado, and S. Hosseini, “Strengthening critical thinking in engineering students,” *Int J Interact Des Manuf*, vol. 13, no. 3, pp. 995–1012, Sep. 2019, doi: 10.1007/s12008-019-00559-6.

[29] B. D. Ebright-Jones and K. S. Cortina, "Critical Thinking in College – Can We Teach It?," *Zeitschrift für Pädagogik*, no. 2, pp. 145–164, Mar. 2025, doi: 10.3262/zp2502145.

Appendix

Table 1: Critical thinking dimensions defined (Facione and colleagues)

Interpretation: The ability to comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures, or criteria.	**Analysis**: The ability to identify the intended and actual inferential relationships among statements, questions, concepts, descriptions, or other forms of representation intended to express beliefs, judgments, experiences, reasons, information, or opinions.	**Evaluation**: The ability to assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief, or opinion, and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions, or other forms of representation.
Inference: The ability to identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses; to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation.	**Explanation**: The ability to state the results of one's reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological, and contextual considerations upon which one's results were based; and to present one's reasoning in the form of cogent arguments.	**Self-Regulation**: The ability to self-consciously monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgments with a view toward questioning, confirming, validating, or correcting either one's reasoning or one's results

Table 2: Single factor ANOVA

Sources	*SS*	*df*	*MS*	*F*	*P value*
Between Groups	0.176	5	0.035	1.397	0.247
Within Groups	0.956	38	0.025		
Total	1.132	43	0.026		