2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #52660

Students' Critical Thinking Conceptions as Predictors of Skill Development in Computer Engineering Capstone Courses

Dr. Trini Sofia Balart, Texas A&M University

Trini Balart completed her PhD in Multidisciplinary Engineering with a focus in engineering education and the impact of AI on education at Texas A&M University, and her Bachelors of Science in Computer Science engineering from Pontifical Catholic University of Chile. Her main research interests include Improving engineering students' learning, innovative ways of teaching and learning, and how artificial intelligence can be used in education in a creative and ethical way.

Dr. Gibin Raju, Texas A&M University

Gibin Raju is a Postdoctoral Researcher in Engineering Education in the Department of Multidisciplinary Engineering at Texas A&M University. He earned his Ph.D. in Engineering Education from the University of Cincinnati and is a recipient of the ASEE Educational Research and Methods (ERM) Division Apprentice Faculty Grant. His research focuses on transformative learning, creativity, critical thinking skills, spatial skills, and engineering design, with particular attention to cognitive stress, cognitive load, and STEM accessibility. By integrating these strategies into a holistic framework, Dr. Raju enhances workforce development, strengthens engineering pathways, and broadens participation in STEM fields. His work equips educators with tools to personalize instruction and foster professional identity, ultimately preparing future engineers to thrive in a rapidly evolving global landscape.

Dr. Kristi J. Shryock, Texas A&M University

Dr. Kristi J. Shryock is the Frank and Jean Raymond Foundation Inc. Endowed Associate Professor in Multidisciplinary Engineering and Affiliated Faculty in Aerospace Engineering in the College of Engineering at Texas A&M University. She also serves as Director of the Craig and Galen Brown Engineering Honors Program. She received her BS, MS, and PhD from the College of Engineering at Texas A&M. Kristi works to improve the undergraduate engineering experience through evaluating preparation in areas, such as mathematics and physics, evaluating engineering identity and its impact on retention, incorporating non-traditional teaching methods into the classroom, and engaging her students with interactive methods.

©American Society for Engineering Education, 2026

Students' Critical Thinking Conceptions as Predictors of Skill Development in Computer Engineering Capstone Courses

Abstract

Critical thinking (CT) is widely cited as a core outcome of engineering education, yet persistent challenges remain in aligning how CT is conceptualized, taught, and assessed. In particular, limited empirical work has examined whether students' own understandings of CT shape their subsequent skill development in authentic engineering contexts. This study investigates whether computer engineering students' initial conceptions of critical thinking predict measurable CT gains during a project-based capstone course. Using an explanatory sequential mixed-methods design, we combined pre-post performance-based assessment with qualitative analysis of students' written CT definitions. Sixteen capstone students completed the California Critical Thinking Skills Test (CCTST) at the beginning and end of the semester and responded to an open-ended prompt defining critical thinking in their own words. Students' definitions were deductively coded using CT skill categories aligned with the CCTST (analysis, inference, evaluation, interpretation, explanation, induction, deduction, and numeracy), generating individual conceptual profiles. Quantitative analyses revealed modest overall gains in CT, with the strongest and most consistent improvement observed in inference-related skills. Qualitative results showed that students most frequently conceptualized CT as problem decomposition and solution generation (analysis, inference, induction), while explanation and deduction were rarely articulated and numeracy was absent. Neither the breadth of students' CT conceptions nor self-reported confidence predicted CT gains beyond baseline performance. Case-level comparisons illustrated that students may demonstrate measurable growth in CT dimensions they do not explicitly associate with "critical thinking" at baseline. Together, these findings suggest that capstone experiences may support development in specific reasoning operations regardless of students' initial CT framings.

1. Introduction

Critical thinking (CT) is widely positioned as a cornerstone outcome of engineering education, linked to graduates' employability and their capacity to navigate ill-structured, high-stakes problems that demand judgment under uncertainty (Winberg et al., 2020). Yet, despite decades of curricular innovation, CT remains difficult to develop and to provide evidence in ways that are both meaningful for engineering practice and methodologically defensible for research (Ahern et al., 2019; Deo & Hölttä-Otto, 2024). Reviews of CT in engineering education continue to document a persistent gap between the aspiration to cultivate higher-order reasoning and the ways CT is actually enacted, taught, and assessed, often inconsistently, implicitly, or through proxies that privilege procedural problem solving over reflective judgment (Ahern et al., 2019; Cooney et al., 2008).

A central challenge is conceptual: CT is a multifaceted construct spanning cognitive skills and intellectual dispositions. Foundational accounts emphasize purposive, self-regulatory judgment, and a constellation of skills such as analysis, inference, evaluation, interpretation, and explanation (Facione, 1990). Complementary traditions emphasize the application of intellectual standards (e.g., clarity, accuracy, depth, breadth, fairness) to reasoning and argumentation, highlighting metacognitive monitoring as a core component of rigorous thinking (Paul & Elder, 2013). Still other definitions frame CT as “reasonable, reflective thinking” oriented toward deciding what to believe or do, underscoring that CT involves more than technical correctness; it requires justification, warrant, and responsiveness to evidence and counterarguments (Ennis, 2018). Within engineering programs, however, CT is frequently conflated with efficient problem solving, algorithmic execution, or the production of a single “correct” answer, particularly in early and mid-curriculum experiences where tasks are well-structured and grading systems reward speed and accuracy (Claris & Riley, 2012; Cooney et al., 2008). This mismatch between what CT is (as theorized) and what CT becomes (as practiced and rewarded) may shape students’ internalized conceptions of CT, conceptions that, in turn, may influence how students approach complex design and decision-making contexts later in the curriculum.

Capstone design courses are often framed as culminating educational experiences that can make CT visible because they require students to define constraints, justify tradeoffs, evaluate alternatives, and communicate decisions to varied audiences. In these settings, students are typically pushed toward reflective work practices, especially when instruction integrates structured reflection, critique, and iterative justification (Konrad et al., 2018). Engineering design tasks can also foreground CT in a discipline-authentic way by requiring argumentation about evidence, requirements, and implications of decisions, activities that are difficult to reduce to procedural templates (Mulligan et al., 2018). Yet, even in capstone contexts, the field has limited empirical evidence clarifying which students develop CT most strongly, why, and how their initial understandings of CT shape their learning trajectories. In other words, capstone may provide conditions that invite CT, but students’ uptake of those conditions may depend on the conceptual frames they bring into the experience.

A related challenge is measurement. CT assessment in engineering education spans rubrics, surveys, standardized CT tests, and customized instruments, each with tradeoffs in validity, feasibility, and fidelity to authentic engineering reasoning (Deo & Hölttä-Otto, 2024). Standardized instruments such as the California Critical Thinking Skills Test (CCTST) are attractive because they provide validated, norm-referenced measures and subscale reporting that can support pre-post comparisons and trend detection (Facione et al., 2008; Liu et al., 2014). At the same time, standardized tests are often critiqued for potentially underrepresenting discipline-specific reasoning and for relying heavily on selected-response items that may not capture the full complexity of design judgment (Deo & Hölttä-Otto, 2024). This tension suggests a methodological opportunity: rather than treating conceptual understanding (what students think CT is) and performance outcomes (what students demonstrate) as separate lines of evidence, researchers can design integrative approaches that connect students’ expressed conceptions to measured skill development.

This study advances an integrative approach by examining whether engineering students’ pre-existing conceptions of CT predict their demonstrated gains in CT skills during a project-based capstone course. Grounded in the CT skills articulated in Facione’s framework (Facione, 1990), the study combines (a) pre-post CCTST measurement and (b) qualitative analysis of open-ended prompts eliciting students’ definitions and conceptualizations of CT.

Specifically, students' written responses are coded using Facione-informed categories and intellectual standards to characterize conceptual profiles (Facione, 1990; Paul & Elder, 2013). These conception categories are then modeled as predictors of CCTST score change using hierarchical regression, enabling an empirical test of whether "what students believe CT entails" is associated with "how much CT they develop" within capstone. A qualitative comparative analysis of deductively coded student definitions provides explanatory depth by contrasting conceptual profiles associated with higher versus lower CT gains, offering insight into the kinds of meanings students attach to CT and how these meanings align or fail to align with measurable development.

By connecting conceptual and performance-based measures, this work addresses a persistent gap in engineering education research: understanding the relationship between students' conceptual starting points and their measurable learning outcomes in complex, culminating experiences (Ahern et al., 2019; Deo & Hölttä-Otto, 2024). Practically, the findings can inform capstone design by identifying conceptual misconceptions or narrow frames that may limit students' engagement with reflective judgment and by motivating targeted support for students who enter capstone with procedure-centric CT conceptions. Methodologically, the study offers guidance for combining validated CT testing with qualitative conceptual profiling, an approach that can improve interpretability of assessment results and strengthen causal reasoning about why CT gains vary across students (Facione et al., 2008; Liu et al., 2014).

This paper contributes to engineering education by a) providing empirical evidence of CT gains and subscale patterns within a computer engineering capstone and b) demonstrating a mixed-methods approach that connects students' expressed CT conceptions to performance-based outcomes. By situating students' conceptions alongside prior work documenting narrow, procedure-focused CT framings among first-year engineering students, this study positions capstone as a later point in a developmental trajectory and examines whether and how students' view of CT evolve by the end of the curriculum.

Research questions

RQ1 (Conceptual profiles): What conceptions of CT do students hold at the conclusion of a capstone course?

RQ2 (Development): What changes are observed in CCTST total and subscale scores pre-post?

RQ3 (Prediction): Do conception categories predict CT gains (ΔCCTST), beyond baseline CT and key covariates?

RQ4 (Explanation): How do high-gain vs low-gain students differ in CT conceptions (and how do these patterns explain the quantitative results)?

2. Theoretical and Empirical Background

2.1 Defining critical thinking for engineering education

In engineering education, CT is commonly framed as a set of higher-order cognitive skills required for reflective judgment and evidence-based decision-making, capacities that align with accreditation expectations for rigorous analysis, reasoned conclusions, and responsible professional practice (Ahern et al., 2019; P. Facione, 1990). Engineering education has increasingly recognized the need to cultivate CT skills, emphasizing the ability to analyze, evaluate, and solve complex problems (Ahern et al., 2019; Deo & Hölttä-Otto, 2024).

A widely used foundation is Facione's Delphi consensus, which conceptualizes CT as purposeful, self-regulatory judgment enacted through core cognitive skills (e.g., interpretation, analysis, evaluation, inference, explanation, self-regulation) and supported by dispositional commitments (e.g., truth-seeking, open-mindedness) (Facione, 1990; Facione, 2008). This consensus remains central in higher-education CT assessment and instruction because it provides an operational language for both measurement and pedagogy (Facione, 1990)

Complementary traditions emphasize (a) CT as "reasonable reflective thinking focused on deciding what to believe or do," highlighting clarity, justification, and warranted conclusions (Ennis, 1985), and (b) intellectual standards (e.g., clarity, accuracy, precision, relevance, depth, breadth, logic, significance, fairness) that can be used to evaluate the quality of reasoning in students' written and verbal explanations (Paul & Elder, 2014). Together, these perspectives provide a coherent basis for distinguishing procedural problem-solving from reflective judgment and for coding students' conceptual accounts of what CT entails.

2.2 Critical thinking in engineering education: persistent challenges

Despite long-standing calls to develop CT, engineering students often encounter curricula and assessment structures that reward procedural fluency and convergent problem solving more than reflective judgment under uncertainty. In a synthesis of CT in engineering education, Ahern et al. (2019) show that definitions and operationalizations of CT vary substantially across studies and that instructional approaches and assessment choices frequently lack alignment, making it difficult to compare outcomes or accumulate evidence across contexts (Deo & Hölttä-Otto, 2024).

At the same time, reviews of CT in engineering and technology education note that CT is frequently invoked as an essential outcome while remaining inconsistently defined and unevenly taught, with common instructional patterns including open-ended tasks and reflective writing that require students to justify decisions rather than simply compute answers (Cooney et al., 2008). This mismatch between CT as an aspirational goal and students' narrower, procedure-centered conceptions motivates a closer examination of the conceptual starting points students bring into advanced engineering learning environments.

2.3 Why capstone is a plausible site for CT growth

Evidence from earlier stages of the curriculum indicates that students' understandings of CT may already be constrained before they reach capstone. A recent study by White et al. (2025) found that many students conceptualized CT primarily as routine problem-solving and task completion with limited attention to evaluative, reflective, or justification-oriented aspects of CT. These early misconceptions highlight the need to examine whether and how CT conceptions have shifted by the time students enter capstone and whether capstone experiences can broaden or reframe these initial understandings.

Capstone design courses are frequently positioned as culminating experiences in which students integrate technical knowledge with real-world constraints, teamwork, and iterative decision-making. These features create conditions where CT should be both needed and observable: students must interpret ambiguous requirements, evaluate alternatives, justify tradeoffs, and defend conclusions with evidence. Evidence from capstone-focused research suggests that pedagogical structure (e.g., the kinds of coaching, feedback, deliverables, and industry interaction students experience) can shape students' perceived skill development and performance-relevant behaviors (Konrad et al., 2018).

Capstone is also closely aligned with broader project-based learning (PjBL) arguments in engineering education: authentic projects can increase opportunities for applying reasoning standards, engaging uncertainty, and practicing justification in context (Uziak, 2016). This type of course is characterized by collaborative problem-solving, iterative design processes, peer feedback, and continuous reflection (Kong et al., 2024). However, even when capstone environments are rich in CT demands, students may respond differently depending on how they understand CT in the first place, whether as "getting the right answer efficiently" or as "making warranted judgments using evidence and standards."

2.4 Measuring CT: performance-based instruments and conceptual measures

A persistent methodological challenge is that CT is often assessed either through performance-based measures or through self-report and course artifacts, with limited integration between students' conceptions of CT and their demonstrated abilities. Large-scale assessment reviews in higher education emphasize the importance of construct clarity, validity evidence, and cautious interpretation when using standardized CT measures, particularly when translating scores into educational claims (Liu et al., 2014).

Within this measurement landscape, the CCTST is a widely used performance measure grounded in the Delphi framework designed to capture core reasoning skills relevant to reflective decision-making (Facione et al., 2008). Yet contemporary engineering-focused reviews also highlight limitations common to CT assessment in the field, such as construct fragmentation across instruments, inconsistent reporting practices, and challenges linking scores to instructional mechanisms (Deo & Hölttä-Otto, 2024). These concerns reinforce the value of pairing performance measures with complementary qualitative evidence that can illuminate how students conceptualize CT and how those conceptions relate to learning trajectories.

3. Methods

This study uses an explanatory sequential mixed-methods approach in a project-based computer engineering capstone course, combining pre-post performance-based assessment with qualitative evidence of students' conceptualizations of CT. The study was conducted in a one-semester, team-based computer engineering capstone course in which students worked on industry-inspired projects, during fall 2025 semester. All procedures followed institutional requirements for human-subjects research, including informed consent and confidentiality protections. The analytic sample consists of students who provided matched pre- and post-assessment data on the CCTST and completed an open-ended prompt defining CT, n=16. In addition, students reported a single-item self-rating of confidence in applying CT to engineering problems (5-point scale) collected alongside the open-ended prompt.

3.1 Instruments

The CCTST served as the primary quantitative measure of CT skills. The CCTST consists of 38 scenario-based multiple-choice items that do not require prior disciplinary knowledge, supporting interpretation as a measure of general reasoning rather than content recall (P. Facione, 1990; P. A. Facione, 2011). The assessment provides reporting across eight dimensions: analysis, interpretation, evaluation, inference, explanation, induction, deduction, and numeracy (Facione et al., 2008).

To capture students' conceptions of CT, students responded to an open-ended prompt eliciting how they define CT and what CT "looks like" in engineering work (for example, how they know they are thinking critically; what counts as good reasoning). Students provided a written definition of CT in their own words. These responses were analyzed using directed qualitative content analysis, a deductive approach appropriate when prior theory guides the initial coding categories. Initial categories were defined a priori using CT skill dimensions aligned with the CCTST and supported by established CT frameworks (e.g., analysis, interpretation, evaluation, inference, explanation, induction, deduction, numeracy). Responses could receive multiple codes when they explicitly or implicitly referenced more than one dimension. Inter-rater agreement was examined on an initial subset of responses and refined through discussion; the remaining data were coded independently with discrepancies resolved through consensus.

3.2 Analysis

CT gains were computed as pre-to-post change scores for the CCTST overall score and, where applicable, for subscale scores. The primary inferential analysis tests whether students' coded CT conception profiles predict CT skill gains (ΔCCTST), consistent with the study's aim of linking what students identify as CT to their measured CT development.

Qualitative analysis focused on how students framed CT at baseline using directed qualitative content analysis grounded in Facione's CT skills. Because the purpose was to characterize students' definitions using a priori CT dimensions rather than to inductively generate new themes, coding proceeded by applying predefined categories to each response while documenting brief rationales grounded in participants' language.

Paired-samples t-tests were used to examine pre-post changes in the CCTST overall score and subscales, with Holm adjustments applied to control the familywise error rate across multiple subscale comparisons. Because the matched sample was small (n = 16), these within-subject inferential tests should be interpreted with caution: estimates may be unstable, statistical power is limited, and assumption violations (particularly departures from normality in change scores) may have a larger influence on results in a sample of this size. Accordingly, p-values are treated as exploratory descriptive indicators and are interpreted alongside effect sizes rather than as strong evidence of generalizable effects.

4. Results

A total of *n* = 16 computer engineering capstone students provided matched pre-post CCTST data and an open-ended definition of critical thinking. Students completed the CCTST pre-post, and the dataset includes subscale scores aligned with the eight CT skill categories used in our qualitative coding (Analysis, Interpretation, Inference, Evaluation, Explanation, Induction, Deduction, Numeracy). Open-ended CT definitions were coded into these categories using the revised coding table and brief rationales.

4.1 Quantitative change in critical thinking

Addressing RQ2 (Development), students completed the CCTST at the beginning and end of the capstone course, allowing examination of within-student change across the overall score and validated subscales (Facione, 1990; Facione et al., 2008). As shown in Table 1, pre-course CCTST overall scores averaged 85.44 (*SD* = 4.97), and post-course scores averaged 87.13 (*SD* = 4.77). Across the matched pre-post sample used for this paper, mean CCTST overall performance increased from pre-test to post-test (ΔM = +1.69). Subscale patterns suggested the largest average gains in Inference (ΔM = +3.00) and Numeracy (ΔM = +3.94), followed by Deduction (ΔM = +2.69). Paired-samples t-tests suggested positive pre-post change for Overall, Inference, Deduction, and Numeracy in the uncorrected analyses. After Holm adjustment for multiple comparisons, Inference was the only subscale that retained this signal. These results are interpreted as exploratory and are reported alongside effect sizes (Cohen's d_z) to support practical interpretation.

Table 1. Pre-Post CCTST Descriptives and Change

Measure	Pre M (*SD*)	Post M (*SD*)	ΔM	d_z
CCTST Overall	85.44 (4.97)	87.13 (4.77)	1.69	0.58
Analysis	88.44 (5.97)	89.50 (6.00)	1.06	0.2
Inference	87.63 (4.76)	90.63 (4.63)	3	0.83
Evaluation	80.81 (8.12)	83.25 (8.08)	2.44	0.25
Induction	85.19 (4.68)	86.81 (4.65)	1.63	0.44
Deduction	85.88 (6.27)	88.56 (6.85)	2.69	0.61
Interpretation	87.44 (6.35)	89.75 (6.85)	2.31	0.42

Explanation	86.75 (4.55)	88.00 (6.53)	1.25	0.2
Numeracy	85.00 (4.95)	88.94 (6.27)	3.94	0.73

4.2. Self-reported confidence applying critical thinking

Addressing RQ3 (Prediction) and RQ2 (Development), students reported high confidence in applying CT to engineering problems (M = 4.38, *SD* = 0.62 on a 1-5 scale; 7 students selected "5", 8 selected "4", 1 selected "3"). Confidence was not associated with CCTST gains ($r = -0.00$, $p = .986$) and was weakly, non-significantly related to pre-course CCTST performance ($r = 0.27$, $p = .315$).

4.3. Students' conceptions of critical thinking

Addressing RQ1 (Conceptual profiles), students' responses to an open-ended prompt ("In your own words, define critical thinking") were coded using CCTST-aligned CT skill categories (analysis, interpretation, inference, evaluation, explanation, induction, deduction, numeracy) to characterize their conceptions of CT. Coding was conducted at the level of whether each response explicitly or implicitly referenced a given CT skill, resulting in a binary profile for each participant.

Students' definitions most commonly emphasized Analysis and Inference (each 12/16), followed by Induction (11/16) and Evaluation (9/16). Interpretation appeared in 7/16 responses, while Explanation (3/16) and Deduction (3/16) were comparatively rare. Numeracy did not appear in any definition (0/16). These patterns mirror students' language and the documented coding rationales, suggesting that many students conceptualized CT primarily as procedural problem-solving rather than as a broader construct involving justification, standards of reasoning, and disciplined evaluation (Facione, 1990; Paul & Elder, 2013).

Table 2. Frequency of CT-skill references in students' definitions

Dimension	n (%)
Analysis	12 (75%)
Inference	12 (75%)
Induction	11 (69%)
Evaluation	9 (56%)
Interpretation	7 (44%)
Explanation	3 (19%)
Deduction	3 (19%)
Numeracy	0 (0%)

To characterize how expansive students' definitions were, a breadth score was calculated as the number of distinct CT dimensions invoked in each response (range: 0-8). Observed breadth ranged from 2 to 6 dimensions (M = 3.56, *SD* = 1.15), indicating that even relatively brief definitions often incorporated multiple elements. The definitions most commonly centered on analysis, inference, and induction. Exemplar responses illustrated these patterns, including emphases on generating a “unique solution” (Inference/Induction), engaging in “careful analysis” linked to producing “ideas and solutions” (Induction/Analysis), and integrating multiple elements to reach a “logical conclusion” (Analysis/Interpretation/Inference/Deduction).

4.5 Linking conceptions to measured gains: associations and predictive models

Addressing RQ3 (Prediction), the study examined whether students’ conceptions of CT, as captured by the open-ended definition, predicted measured gains in CT over the capstone experience. Bivariate associations did not support the expectation that broader conceptions were associated with larger gains on the CCTST (breadth vs. ΔCCTST overall: $r = -0.04$, $p = .873$). Similarly, the presence or absence of any individual coded dimension showed no detectable relationship with gains (all point-biserial ps > .50; Numeracy excluded due to lack of variance). Pre-course performance exhibited a negative, though non-significant, association with gains ($r = -0.36$, $p = .173$), consistent with a ceiling effect in which higher initial scores left less room for improvement. Hierarchical regression analyses reinforced this pattern that adding breadth and confidence did not meaningfully improve prediction beyond pre-course scores alone (Model 1: $R^2 = 0.128$, $p = .173$; Model 2: $R^2 = 0.133$; Model 3: $R^2 = 0.141$).

4.6 Individual-level profiles

Addressing RQ4, we conducted a comparative analysis to provide additional explanatory depth by examining definitions from students with higher gains (ΔCCTST overall ≥ 3; $n = 7$) and those with lower gains or decreases (ΔCCTST overall ≤ 2; $n = 9$). Students in the high-gain group began the course with lower pre-course scores on average; however, their definitions were not consistently broader than those of lower-gain students. For example, P12 ($\Delta = +6$) explicitly referenced integrating components to reach a “logical conclusion,” indicating the presence of Deduction, whereas P16 ($\Delta = -5$) emphasized understanding the situation and solving problems independently, referencing Analysis/Interpretation/Inference/Induction). These cases illustrate that similarly articulated conceptions of CT can correspond to markedly different learning trajectories during capstone, pointing to the need for further investigation of instruction-specific and task-specific factors that influence demonstrated gains.

4.7 Integration highlight

Taken together, the quantitative and qualitative findings suggest a more selective picture of CT development in capstone than a simple ‘more conceptually sophisticated students improve more’ interpretation. Quantitatively, the clearest within-sample improvement pattern appeared in Inference. Qualitatively, students’ baseline definitions already emphasized analysis, solution generation, and reaching conclusions under uncertainty, indicating that many entered capstone with a conception of CT that partially aligned with inference-heavy design work. This alignment may help explain why inference-related performance showed the strongest improvement pattern.

At the same time, qualitative results also help explain the null predictive findings. Students rarely articulated explanation and deduction, and breadth of definition did not distinguish higher-gain from lower-gain students in a consistent way. Case-level comparisons further suggest that students may develop performance in dimensions they do not initially verbalize, implying that capstone task demands, feedback, and repeated decision-making opportunities may shape growth beyond what is captured in a short baseline definition.

5. Discussion

5.1. Summary of key findings

In response to RQ2, students demonstrated a modest increase in overall CCTST performance across the capstone semester, with the clearest subscale signal in Inference. In response to RQ1, students' incoming conceptions of CT most frequently referenced Analysis, Inference, and Induction, whereas Explanation and Deduction were rarely articulated and Numeracy was absent. For RQ3, neither definitional breadth nor the single-item confidence rating predicted overall ΔCCTST, and regression results indicated minimal incremental explanatory power of these predictors beyond baseline scores in this small sample. Finally, RQ4 case-level comparisons suggested that students may demonstrate gains in performance dimensions they do not spontaneously associate with "critical thinking" at baseline (e.g., numeracy or evaluative judgment), highlighting the value of case-level analyses when sample sizes are small.

5.2. Interpretation through Facione / CT theory

The quantitative pattern, showing strongest gains in Inference, aligns with the types of reasoning moves routinely required in capstone courses, including drawing warranted conclusions from incomplete information, selecting among plausible alternatives, and updating decisions as constraints and stakeholder requirements evolve. From a Facione-aligned perspective, inference is not simply "guessing". It involves forming and testing hypotheses, drawing conclusions from available evidence, and judging what follows from information under uncertainty (Facione, 2011). In a project-based capstone setting, students frequently encounter ill-structured problems with multiple feasible pathways, making inference-heavy reasoning a natural target for practice and feedback (Konrad et al., 2018). This alignment between capstone task structure and inference-related performance may help explain why inference showed the most consistent growth. The contrast between modest overall gains and relatively stronger inference gains is therefore consistent with a picture in which capstone selectively reinforces specific reasoning operations that are heavily exercised by course tasks, rather than uniformly elevating all CT dimensions (Konrad et al., 2018).

At baseline, many students described CT as a process of breaking down a problem and generating a solution, using language consistent with Analysis/Inference/Induction. This framing is pedagogically valuable, as it reflects authentic engagement with open-ended problems; however, it may also indicate a narrower, procedure-forward conception that underemphasizes justification and evaluative standards. The relative absence of Explanation and Deduction in students' definitions suggests that many did not spontaneously frame CT as making reasoning explicit , articulating why a conclusion follows, or distinguishing formal validity from plausible inference. In professional engineering practice, however, reasoning quality is often evaluated through justification, defensible tradeoffs, and the logic connecting evidence to claims (Francis et al., 2022). Thus, even when students recognize CT as "solving open-ended problems," they may not explicitly foreground the intellectual standards by which solutions are judged, particularly standards related to explanation, validity, and evaluation.

When viewed alongside findings from first-year engineering populations, a developmental pattern begins to emerge. Recent qualitative work with first-year engineering students similarly found that many students equate CT with routine problem-solving and task completion, with relatively few references to evaluative, reflective, or justification-oriented aspects of CT (White et al., 2025). Together, these results suggest that procedure-centric CT framings can arise early in the curriculum and persist into capstone, even as certain reasoning skills (such as inference) show measurable growth.

Several factors may help explain why students' conceptions of critical thinking did not predict measured gains. First, baseline CCTST scores exhibited a ceiling-like pattern, with higher pre-scores associated with smaller gains, thereby constraining variance in Δ scores. Second, conceptions were assessed using a single short open-ended prompt; students may hold richer, more tacit understandings of CT than they articulate in a brief definition, and brief responses can compress meaningful differences into the same coding categories. Third, the coding scheme, intentionally aligned with CCTST reporting dimensions, captures whether a definition references a given dimension but not the depth or quality of reasoning about that dimension (e.g., merely mentioning "evaluate" versus articulating a principled evaluative stance). Finally, capstone experiences may exert strong situational effects, including team practices, feedback cycles, sponsor expectations, and milestone structures, that shape performance growth in ways not directly traceable to baseline definitional language, potentially overshadowing individual differences in baseline definitions.

The individual profiles help interpret these issues. Some high-gain students showed substantial improvement in subscales that were not present in their definitions (e.g., numeracy or evaluation), suggesting that capstone may develop performance dimensions that students do not initially associate with "critical thinking." Conversely, some low-gain students articulated multi-dimensional conceptions but did not show measurable gains, consistent with the possibility that articulation and performance can diverge, particularly when performance is sensitive to test conditions, motivation, or the specificity of instruction. These case-level patterns suggest that baseline conceptions may be better treated as a diagnostic window into students' language and framing rather than as a deterministic predictor of performance change.

5.3. Methodological implications

Methodologically, these findings support the value of integrating conceptual measures (what students say CT is) with performance-based measures (what students can demonstrate). Even when predictive relationships are weak, this integration surfaces actionable mismatches: dimensions that students rarely name (e.g., explanation and deduction) may be precisely the dimensions instructors need to make visible and practice explicitly. The approach used here, deductive coding aligned to CCTST reporting dimensions, offers a replicable strategy for "quantitizing" a minimal qualitative dataset in a way that supports mixed-methods integration. In contexts where qualitative data collection is necessarily light, such as brief surveys embedded in a course, a case-ordered joint display and alignment-based interpretation can provide explanatory value without overstating statistical generalizability.

For capstone design, the findings suggest two complementary instructional moves. First, evaluative and justificatory reasoning should be made explicit through prompts and rubrics that require students to state assumptions, defend tradeoffs with evidence, and articulate why a conclusion follows from constraints and data. Second, students' incoming conceptions should be treated as diagnostic rather than deterministic. A short baseline prompt can identify CT dimensions that are missing or underdeveloped in students' language (e.g., explanation and validity), enabling targeted scaffolding early in the semester through tools such as decision memos, structured argumentation, or critique protocols that emphasize explanation and evaluation.

5.4. Limitations

This study is limited by the single-course, single-institution context and the small matched sample ($n = 16$), which constrains statistical power and increases sensitivity to outliers. Although the CCTST provides a standardized indicator of performance, it does not capture all forms of CT that are relevant in capstone contexts, such as collaborative reasoning, domain-specific argumentation, or sustained justification across a semester-long design history. The qualitative dataset was intentionally minimal, relying on a single open-ended definition prompt, and the coding framework emphasized alignment with CCTST dimensions; as a result, other epistemic framings of CT (e.g., ethical reasoning, professional judgment, social negotiation in teams) may be underrepresented. Additionally, the single-item confidence measure is a coarse indicator and should be interpreted as a contextual indicator rather than a robust self-efficacy construct. Finally, the analytic sample includes only students who completed all pre-post measures and the open-ended prompt, which may introduce selection bias toward more engaged or higher-performing students and further limits generalizability.

5.5. Future work

Future studies could extend this design across multiple courses and institutions to test the robustness of the inference gain pattern and to examine whether specific instructional features, such as feedback cycles, decision documentation, design review practices, moderate CT development. Building on prior work documenting narrow, task-focused CT conceptions among first-year engineering students, a multi-cohort or longitudinal design that uses common conceptual prompts and CT frameworks at both the introductory and capstone levels could clarify when and how students' framings of CT shift across the curriculum. Methodologically, richer conception measures, including multiple prompts, scenario-based concept questions, or rubric-scored written explanations, could capture qualitative distinctions beyond presence/absence coding. Longitudinal designs spanning multiple curricular stages would further enable examination of whether targeted reframing efforts early in the curriculum reduce procedure-centric misconceptions by the time students reach capstone. A particularly promising direction is to connect incoming conceptions to capstone artifacts (e.g., design reports, decision logs, peer critique) to examine how conceptual framing translates into observable reasoning practices over time. Finally, intervention studies that explicitly target conceptual reframing by teaching students to define CT in terms of evidence-based justification, evaluation, and validity could test whether strengthening conceptions leads to stronger performance growth.

6. Conclusion

This study examined whether computer engineering capstone students' conceptions of CT predict measurable CT development during a project-based capstone experience. Using matched pre-post CCTST data (n = 16), we observed modest overall growth in CT performance across the semester with the clearest observed improvement in Inference and Numeracy. At baseline, students' open-ended definitions most frequently emphasized problem decomposition and solution generation under uncertainty (Analysis, Inference, and Induction), while fewer students explicitly articulated CT as justifying decisions (Explanation) or distinguishing the validity of conclusions through formal logic (Deduction). Neither definitional breadth nor a single-item confidence rating meaningfully predicted overall ΔCCTST in this sample.

Taken together, the findings suggest that capstone experiences may support gains in specific reasoning operations, particularly inference-related judgment, while students' initial language about CT remains more strongly anchored in analysis and solution generation. Methodologically, this study demonstrates a mixed-methods approach for connecting a brief conceptual prompt to a standardized performance measure by coding definitions into CCTST-aligned dimensions and integrating case-level profiles with pre-post outcomes. Practically, instructors can use short baseline conception prompts as diagnostics to identify under-articulated CT dimensions (e.g., explanation and deduction) and to design early scaffolds that require explicit justification, evidence-based tradeoffs, and defensible reasoning in team deliverables.

Given the single-course context, small matched sample, and reliance on a single open-ended prompt, these results should be interpreted as exploratory. Future work should replicate this design across multiple capstone settings, use richer measures of CT conceptions (e.g., multiple prompts or scenario-based items), and connect conceptual profiles to capstone artifacts (e.g., decision memos and design reviews) to better characterize how students' framings of CT translate into sustained reasoning practices over time.

References

Ahern, A., Dominguez, C., McNally, C., O'Sullivan, J. J., & Pedrosa, D. (2019). A literature review of critical thinking in engineering education. *Studies in Higher Education*, *44*(5), 816–828. https://doi.org/10.1080/03075079.2019.1586325

Claris, L., & Riley, D. (2012). Situation critical: Critical theory and critical thinking in engineering education. *Engineering Studies*, *4*(2), 101–120. https://doi.org/10.1080/19378629.2011.649920

Cooney, E., Alfrey, K., & Owens, S. (2008). Critical thinking in engineering and technology education: A review. *2008 Annual Conference & Exposition*, 13.344. 1-13.344. 16.

Deo, S., & Hölttä-Otto, K. (2024). Critical thinking assessment in engineering education: a scopus-based literature review. *Journal of Mechanical Design*, *146*(7), 072301. https://doi.org/10.1115/1.4064275

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, *43*(2), 44–48.

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, *37*(1), 165–184. https://doi.org/10.1007/s11245-016-9401-4

Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). In *Educational Resources Information Center (ERIC)* (pp. 1–112).

Facione, P. A. (2011). Critical thinking: What it is and why it counts. *Insight Assessment*, *1*(1), 1–23.

Facione, P. A., Facione, N. C., Blohm, S., & Gittens, C. A. (2008). *California critical thinking skills test manual. CA: Insight Assessment*. The California Academic Press.

Francis, R. A., Paretti, M. C., & Riedner, R. (2022). Theorizing engineering judgment at the intersection of decision-making and identity. *Studies in Engineering Education*, *3*(1), 79. https://doi.org/10.21061/see.90

Kong, S.-C., Cheung, M.-Y. W., & Tsang, O. (2024). Developing an artificial intelligence literacy framework: Evaluation of a literacy course for senior secondary students using a project-based learning approach. *Computers and Education: Artificial Intelligence*, *6*, 100214. https://doi.org/10.1016/j.caeai.2024.100214

Konrad, R., Hall-Phillips, A., & Vila-Parrish, A. R. (2018). Are our students prepared? the impact of capstone design pedagogical approaches on student skill development during industry-sponsored fieldwork. *INFORMS Transactions on Education*, *18*(3), 183–193. https://doi.org/10.1287/ited.2018.0198

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, *2014*(1), 1–23. https://doi.org/10.1002/ets2.12009

Mulligan, R. P., Simper, N., & Mulligan, N. (2018). An engineering design course to develop and assess critical thinking and problem solving. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.v0i0.13084

Paul, R., & Elder, L. (2013). *Critical thinking: Tools for taking charge of your professional and personal life*. Pearson Education.

Uziak, J. (2016). A project-based learning approach in an engineering curriculum. *Global Journal of Engineering Education*, *18*(2), 119–123.

Winberg, C., Bramhall, M., Greenfield, D., Johnson, P., Rowlett, P., Lewis, O., Waldock, J., & Wolff, K. (2020). Developing employability in engineering education: A systematic review of the literature. *European Journal of Engineering Education*, *45*(2), 165–180. https://doi.org/10.1080/03043797.2018.1534086

White, L. L. A., Raju, G., Watson, K., & Shryock, K. J. (2025, June 22). Critical thinking (Mis)conceptions of first-year engineering students. ASEE. https://peer.asee.org/critical-thinking-mis-conceptions-of-first-year-engineering-students