© 2026 Tunç. This article follows the Open Access policy of CC BY NC under CCA license v 4.0.

Submitted: 28/04/2026 - Accepted: 26/05/2026 - Published: 28/06/2026

Professional Judgment in Vocational Education: A Scenario-Based Study of Critical Thinking in Associate-Degree Accounting Trainees

Cevdet Tunç

Bolu Abant Izzet Baysal University, Bolu, Turkey
cevdettunc@ibu.edu.tr

DOI: 10.26417/88drjs57

Abstract

This study aimed to examine the critical thinking and problem-solving skills of associate-degree accounting students through realistic professional scenarios. The research is based on a qualitative design. The participants were fourteen students enrolled in the second year of the accounting program at a vocational college in Turkey's Western Black Sea Region. Data were collected through scenario-based interviews consisting of three professional scenarios and probing questions corresponding to the six dimensions of Facione's taxonomy. Responses were scored by two independent coders using a rating scale; inter-coder reliability was calculated using the Cohen's Kappa coefficient, and the data were also analyzed using thematic analysis. The findings revealed that nine participants fell into the "deep thinker" profile, while five were classified as "developing thinkers." The highest average score was observed in the Analysis dimension, while the lowest average was found in the Interpretation dimension. The overall average was found to be 2.45 out of 3. This pattern suggests that students are relatively strong in solving a problem and identifying its causes, but weaker in correctly defining the problem from the outset. The lowest average across the scenarios was observed in the software and document mismatch scenario, which required root cause analysis. The thematic analysis identified five themes: evidence-based reasoning, cumulative risk perception, self-correction with expert support, internal control, and professional responsibility. The results indicate that realistic and field-specific scenarios make associate-degree students' reasoning skills more apparent than context-free tests.

Keywords: professional skepticism, expert–novice differences, vignette method, higher-order thinking, qualitative case study, Facione taxonomy

Introduction

Vocational and technical higher education prepares students for direct entry into the workforce; associate-degree programs in particular compress occupational training into two years and move graduates into professional roles soon after completion. The value of these programs—both as a route of social mobility and as a source of skilled practitioners—depends not only on the technical content they transmit but also on the higher-order reasoning their graduates can apply in unfamiliar situations (Nie & Mastor, 2024). Accordingly, vocational education policy has gradually shifted from the coverage of technical curricula toward transferable competencies such as critical thinking and problem-solving (Romero-Díaz de la Guardia et al., 2024; Vignjević Korotaj & Buchberger, 2026). Because each profession defines the judgment it expects of a newcomer in its own terms (Abbott, 1988), examining how associate-degree students reason within an authentic professional context speaks to a broader concern: how vocational programs translate classroom learning into professional judgment.

Accounting provides a revealing case. The profession is evolving from a field based on technical knowledge into one focused on reasoning and decision-making: as routine recording and reporting tasks are automated, accountants are expected not only to record transactions accurately but also to apply critical thinking in uncertain, unstructured situations. This has brought critical thinking and problem-solving skills into the core objectives of accounting education, and the mismatch between the skills graduates acquire and those employers seek further strengthens interest in the field (Jackling & De Lange, 2009; Gyekye & Amo, 2024).

The development and assessment of critical thinking and problem-solving skills is actively discussed in the accounting education literature (Wolcott & Sargent, 2021), yet most studies focus on four-year undergraduate programs using quantitative techniques that lack contextual relevance (Sinnewe et al., 2023). Students at two-year associate-degree vocational colleges are relatively underrepresented, even though they enter the workforce immediately upon graduation and face real-world accounting challenges soon after (Nie & Mastor, 2024)—conditions that warrant empirical investigation of their thinking skills.

A second gap concerns measurement. Critical thinking has mostly been assessed through multiple-choice tests, while how students reason when facing a real-life situation has been relatively under-examined, and qualitative studies using realistic professional scenarios remain limited—particularly, in the Turkish context, with associate-degree accounting students. This study addresses all three gaps.

This study aims to elicit the critical thinking and problem-solving skills of associate-degree accounting students through realistic professional scenarios, using Facione's (1990) six-dimensional framework—increasingly a foundational tool in the accounting education literature (Terblanche & De Clercq, 2021; Turner & Tyler, 2023)—as a scoring rubric. Data were collected through scenario-based interviews and analyzed using thematic analysis. The study sought to answer the following questions:

1. What is the distribution of critical thinking profiles among associate-degree accounting students as revealed by scenario-based interviews?
2. How do students' critical thinking scores differ across the six dimensions of Facione's taxonomy?
3. How do students' scores differ across the three scenarios?
4. What themes emerge from students' reasoning regarding accounting problems?

Literature Review

The role of critical thinking in accounting education has a long research history. Kimmel (1995) proposed an early framework for integrating it into accounting courses, while Baril et al. (1998) found that the profession defines critical thinking through attitudes as well as cognitive skills, with practitioners expecting new hires to identify a problem before solving it. More recent work has translated this into structured competency frameworks: Terblanche and De Clercq (2021) developed one specific to accounting students based on Facione's (1990) six dimensions, and Turner and Tyler (2023) applied it in an accounting course, showing the dimensions could be observed even at an introductory level with purposeful instructional design. These two studies are the closest precursors to the present research.

However, the field still lacks a consensus definition. Wolcott and Sargent (2021) note the absence of a single agreed-upon definition of critical thinking in accounting education and call for more rigorous empirical work. Through content analysis of four professional organizations' competency frameworks, Terblanche et al. (2023) found that the term "critical thinking" appeared in only one, with proxy terms such as judgment, decision-making, and problem-solving predominating, and that the self-regulation and inference dimensions were almost entirely absent. This definitional ambiguity thus extends beyond the academic literature to the profession's own regulatory texts.

This trend also appears in teaching practice. Dickins and Reid (2023) found that small changes based on Bloom's taxonomy improved students' ability to apply concepts, though students often faced questions above their developmental level. Hawk (2025) reported that instructors emphasized application and analysis over evaluation and creation, while Sinnewe et al. (2023) identified resource limitations, subject density, and student engagement as the main impediments. Chabrak and Craig (2013) observed that many students assessing the Enron case reasoned within the existing

system rather than critically, whereas Papageorgiou (2023) found students benefited from a dedicated critical thinking course. More broadly, Rothinam et al. (2025) concluded that active, interdisciplinary instruction supports critical thinking while overreliance on standardized tests hinders it. These studies suggest the impediments lie not only in limited resources but in technically centered teaching practices.

Problem-solving, though closely related to critical thinking, is treated as a distinct concept. Jones and Davidson (2007) found that students with higher cognitive complexity performed better on unstructured but not structured questions, suggesting problem-solving becomes evident mainly in ambiguous situations. Jackling and De Lange (2009) reported a disconnect between graduates' technical knowledge and the transferable skills employers prioritize, with graduates ranking problem-solving as the second most important skill for career advancement. Asonitou and Hassall (2019) similarly documented a skills gap in Greece, identifying the ability to solve unstructured problems as the most urgent priority. Together these studies indicate that problem-solving is seen as decisive for professional success yet remains insufficiently developed during education.

In accounting and auditing, these skills are reflected in professional skepticism. Hurtt (2010) developed a tool defining the concept through six characteristics: an inquiring mind, suspension of judgment, information seeking, interpersonal understanding, autonomy, and self-confidence. Suspension of judgment and an inquiring mind largely overlap with Facione's evaluation and inference dimensions. Professional skepticism can thus be seen as the domain-specific manifestation of critical thinking in accounting judgment, bridging the critical thinking and accounting judgment literatures.

Although much of this work is Anglo-American, local and cross-cultural studies also exist. Omurgonulsen and Omurgonulsen (2009) examined the İmarbank case and underlined the importance of pairing critical inquiry with ethical sensitivity. In a latent profile analysis of students from Turkey, South Korea, and the United States, Demir et al. (2023) found a similar profile structure across nations but lower scores for Turkish students on all sub-dimensions. Whether this reflects a true skill gap or the instrument's cultural fit remains unclear, pointing to the need for domain-specific, pragmatic measurement.

How critical thinking is measured is as contested as how it is defined, since context-independent tests do not always reveal how a student reasons in a real situation. A scenario-based approach offers a solution. Hainmueller et al. (2015) showed that vignette-based measures replicate real decision patterns whereas single-scenario designs perform poorly, supporting the use of multiple scenarios and probe questions. Poulou (2001) likewise notes that a common stimulus makes responses comparable and that hypothetical framing reduces social desirability. The design principles behind such tools are detailed below.

Evaluated collectively, this literature relies largely on undergraduate accounting programs and quantitative measurement tools. The qualitative examination of critical thinking and problem-solving among associate- degree vocational school students through realistic professional scenarios remains under-explored. Given the few scenario-based qualitative studies with associate-degree accounting students in Turkey, this study seeks to elicit these students' critical thinking and problem-solving skills through scenario-based interviews.

Theoretical Framework

Critical Thinking and the Facione Framework

Critical thinking has been studied for a long time without a single established definition. Ennis (1985) defines it as reflective and logical thinking focused on deciding what to believe or do, while Paul and Elder (2020) view it as an individual's reflection on their own thought process—a form of higher-order self-regulation. The two approaches are not mutually exclusive: the first emphasizes the product of thinking, the second the process. The most common framework bringing them together emerged from a consensus-building effort by the American Philosophical Association.

Reported by Facione (1990), this Delphi study defined critical thinking, through a multi-stage consensus among forty-six experts, as an evidence- and criterion-based, purposeful, self-regulatory judgment process. This definition and its six-dimensional structure were adopted as the primary theoretical framework for three reasons: it links critical thinking to observable cognitive skills rather than an abstract trait; each dimension can be converted into a rubric and scored consistently by two raters; and it covers both cognitive skills and their application to real situations, making it compatible with a scenario-based design.

The Facione framework operationalizes critical thinking across six cognitive dimensions: interpretation, grasping the meaning of a situation and framing the problem correctly; analysis, identifying relationships among pieces of information and distinguishing a problem's components; evaluation, weighing the reliability of claims and evidence against established criteria; inference, reaching reasonable conclusions from available evidence; explanation, presenting a consistent rationale for the conclusion; and self-regulation, reviewing and, when necessary, correcting one's own reasoning.

These six dimensions are not strictly separated; Facione (1990) emphasizes that they operate interactively and iteratively rather than linearly. An accountant resolving a document discrepancy, for instance, may begin with interpretation, move to analysis, and return to interpretation as inference yields new information, with self-regulation encompassing the other five. Critical thinking is therefore not a single skill but a combination of cognitive processes that feed into one another.

The framework also has strong empirical support. Payan-Carreira et al. (2022) showed that a brief scale based on this six-dimensional structure can be measured validly and reliably in a multinational sample, and it is increasingly used in accounting education, where Terblanche and De Clercq (2021) adapted it for accounting students and Turner and Tyler (2023) observed the dimensions in an accounting course. It was therefore adopted as the basis for this study's scoring rubric.

Expert-Novice Difference and Problem Representation

The Facione framework determines which dimensions to measure; however, a second lens from the tradition of cognitive research is needed to interpret why a dimension profile might be uneven. Studies on the expert-novice difference provide this lens. Chi et al. (1981) demonstrated that novices represent a problem through surface features, whereas experts frame it through deep structure and principles. Similarly, Schoenfeld and Herrmann (1982) found that as expertise increases, perception shifts from surface structure toward solution principles. These findings indicate that problem representation changes with the level of expertise.

This lens generates an expectation about the dimensional profile of novice practitioners. Associate-degree students whose experience is limited to internships may be expected to lag in framing the problem (interpretation) relative to the solution-oriented dimensions such as analysis. Any gap between the dimensions is therefore not merely descriptive but a pattern predicted by the expert-novice framework, which serves as an analytical lens for interpreting the study's most prominent finding.

Problem-Solving and Professional Judgment

Problem-solving is treated as a skill that is intertwined with but distinct from critical thinking. Structured problems have a single correct path and a predetermined solution. Problems encountered in professional life, however, are mostly unstructured; they contain incomplete information, are open to multiple possible solutions, and require the decision-maker to reach a judgment. Jones and Davidson (2007) empirically demonstrated that students exhibit different levels of cognitive complexity when dealing with unstructured problems compared to structured ones. From this perspective, problem-solving can be viewed as the application of critical thinking to a real-world professional situation.

The field-specific counterpart of this skill is professional skepticism. Hurtt (2010) defined it as tendencies such as evidence-seeking, an inquisitive attitude, and suspending judgment, some of which overlap with Facione's evaluation and inference dimensions. The study's framework thus has three layers: Facione's six dimensions at the general cognitive level, expert-novice problem representation at the developmental level, and professional skepticism at the field-specific level. This structure makes it possible to gauge how far students show an inquisitive attitude when faced with unstructured accounting problems.

This three-layered framework generates three expectations that guide interpretation of the findings. First, the dimension profile of novices should be uneven, with interpretation lagging behind the solution-oriented dimensions. Second, the structure and concreteness of a scenario should influence which dimension emerges most prominently. Third, early forms of professional skepticism should be evident in students' reasoning. These expectations align with the study's four research questions and frame the analysis.

Scenario-Based Assessment of Thinking Skills

The method used to measure thinking skills directly influences the quality of the findings. Context-free multiple-choice tests measure skills abstractly and do not always reveal how a student reasons in a real-world situation. The scenario-based approach addresses this limitation by embedding the skill within a real professional situation and asking the student to think through it.

Designing a scenario-based tool also requires a theoretical choice. Rungtusanatham et al. (2011) noted that such an instrument should consist of a common section that keeps contextual information constant and a variable section that triggers reasoning, ensuring all participants encounter the same situation while differences in their responses stem from their own reasoning. The scenarios in this study followed a similar logic: after presenting a common professional scenario, each was explored in depth using probing questions corresponding to Facione's dimensions.

Empirical evidence supports scenario-based measurement. Hainmueller et al. (2015) showed that vignette-based measures yield results consistent with actual behavior—suggesting that reasoning about a fictional accounting scenario approach reasoning in real work settings—and Poulou (2001) noted that the technique can reveal thinking processes in educational research. Realism matters most for associate-degree students with internship experience: relating a hypothetical scenario to their own work makes the judgment they display more authentic. The approach therefore provides an appropriate measurement framework for this study.

Method

Research Model

This study adopted a qualitative research approach. The critical thinking and problem-solving skills of associate-degree accounting students were examined through scenarios involving realistic professional situations. Data were collected via scenario-based, semi-structured individual interviews. To elucidate students' reasoning processes, the six-dimensional critical thinking taxonomy proposed by Facione (1990) was adopted as the theoretical framework. These dimensions are interpretation, analysis, inference, evaluation, explanation, and self-regulation.

Participants

The participants in this study are second-year students enrolled in the bookkeeping (Accounting and Tax Practices) program at a vocational college in Turkey's Western Black Sea Region. Participants were selected on a voluntary basis using purposive sampling. The inclusion criteria were being enrolled in the second year of the program and having field experience that would allow professional judgment to emerge clearly during the interviews. Accordingly, fourteen students participated in the study. The participants' ages range from 19 to 23, and the majority are female. Thirteen of the fourteen students had previously completed an internship in the field of accounting. All participants are 18 years of age or older. Demographic characteristics are presented in Table 1.

The sample size was determined consistently to support the in-depth descriptive nature of the qualitative research. Since no new codes or themes emerged in the final interviews during the data collection and analysis process, data saturation was deemed to have been reached, and no additional participants were sought. In this regard, the sample size is related not to the goal of statistical generalization but to the objective of detailing the participants' reasoning processes.

Table 1. *Demographic characteristics of the participants*

Code	Age	GPA	Internship	Career intention
K1	20	2.92	20 days (public sector)	Intends to pursue
K2	21	2.76	23 days (CPA office)	Considering
K3	20	3.05	1 month (CPA office)	Not specified
K4	20	3.00	20 days (CPA office)	Not specified
K5	20	2.90	1 month (private sector)	Intends to pursue
K6	20	3.02	1 year + 4 months (paid)	Intends to pursue
K7	20	52–55/100	20 days (CPA office)	Intends to pursue
K8	23	1.60	1 month (factory)	Undecided
K9	20	1.76	None	Intends to pursue
K10	19	2.00	3 months (paid)	Intends to pursue
K11	21	2.70	20 days (production)	Intends to pursue
K12	21	Medium	21 days (factory)	Does not intend

Code	Age	GPA	Internship	Career intention
K13	21	2.80	2 years (CPA office, paid)	Intends to pursue
K14	20	2.30	23 days (CPA office)	Intends to pursue

Note. GPA reported on a 4.00 scale unless otherwise indicated. CPA = certified public accountant (SMMM) office.

Scenario Development and Validation

The scenarios were developed to reflect realistic, unstructured problems that associate-degree accounting graduates are likely to encounter early in professional practice. Three were selected: discrepancies between accounting software records and source documents; a year-end physical inventory count that cannot be reconciled with the recorded balance; and an accounting error carried over from a previous period. Each has no single predetermined solution and requires the respondent to define the problem, examine its possible causes, and reach a defensible decision—the reasoning process Facione's (1990) framework aims to elicit.

The scenario design followed Rungtusanatham et al. (2011), in which a scenario-based tool consists of a common module that maintains a consistent professional context and a variable module that triggers reasoning. Here, each scenario's narrative served as the common module and its six probing questions as the variable module. Each question corresponded to a Facione dimension, so every scenario addressed interpretation, analysis, inference, evaluation, explanation, and self-regulation in sequence.

Content validity was assessed through expert review: two experts evaluated the scenarios and probing questions for professional realism, clarity, and alignment with the targeted Facione dimensions. Based on their feedback, the scenarios were revised, question wording was adjusted, and the scenarios were adapted to the institution's specific context.

Before the main data collection, a pilot interview was conducted with two students not participating in the main study to verify the comprehensibility of the scenarios, the adequacy of the probing questions, and the approximate interview duration. As the scenarios proved understandable, no changes were made to the questions.

Data Collection Tool and Process

Data were collected using a scenario-based, semi-structured interview form with three scenarios representing authentic professional situations: a software-document discrepancy, a year-end inventory count mismatch, and an accounting error carried over from a previous period. Each scenario was followed by six probing questions corresponding to the six Facione dimensions, designed to elicit students' problem

definition, cause analysis, solution generation, and reflection on their own reasoning. The full form, including complete scenarios and probes, is provided in Appendix A.

The interviews were conducted one-on-one in Turkish and lasted approximately 40-50 minutes. Before each interview, students were told that the session was not a test, would not be graded, and could be withdrawn from at any point, even afterward—conditions intended to help them express their views comfortably.

After obtaining participants' written consent, the interviews were audio-recorded and transcribed verbatim. The transcripts were cross-checked against the recordings for accuracy, anonymized using participant codes (K1–K14), and divided into sections based on the script and questions, with paralinguistic markers such as pauses and hesitations preserved. Leading statements were avoided throughout, and students were given time to think during moments of silence.

Each participant then received their own transcript and confirmed whether their thoughts had been accurately represented (member checking). Excerpts used in the analysis were translated into English by a faculty member independent of the researcher with language proficiency and knowledge of the field's terminology. The researcher also kept a research diary throughout data collection, documenting participants' comfort, openness, and willingness to participate; its entries indicate that no problematic situations arose during the interviews.

Data Analysis

Data analysis was conducted using a two-tiered structure. The first tier consists of structured rubric scoring based on Facione's (1990) framework; the second tier consists of the thematic analysis procedure described by Braun and Clarke (2006). These two tiers combine a dimension-based structured assessment with the discovery of patterns emerging from the data. The scoring rubric is provided in Appendix B.

In the first layer, each response was scored as 1 (low), 2 (medium), or 3 (high) on the relevant Facione dimension, with unanswered or off-topic responses assigned 0. Each probe mapped to a single dimension, so each scenario was scored out of 18 and the three scenarios out of 54. Scoring was based on the content rather than the length of the response. The full rubric, with the indicator criteria used for each level, is provided in Appendix B.

Scoring was conducted by two independent coders: one researcher and one faculty member with publications in qualitative research and coding. Before proceeding with independent scoring, the two coders conducted pilot coding on the first two transcripts and confirmed a shared understanding of the rubric definitions. During the scoring process, each dimension was evaluated independently of the others, and demographic information was excluded from the process to ensure it did not influence the scoring.

Inter-coder reliability was calculated using Cohen's Kappa coefficient (Cohen, 1960) and found to be .78; the weighted Kappa value calculated for the ordinal scale was .83. Dimension-specific reliability values are presented in Table 2. The thirty cells where disagreement was observed between the two coders were evaluated by a third independent reviewer based on the rubric criteria and assigned a single consensus score (Cascio et al., 2019). These consensus scores were used in subsequent analyses.

After scoring was completed, participants were categorized into three thinking profiles based on their total scores out of 54: superficial thinker (18–29 points), developing thinker (30–41 points), and deep thinker (42–54 points). This classification provided a framework for interpreting the qualitative findings.

Table 2. *Inter-coder reliability (Cohen's kappa) by Facione dimension*

Dimension	Exact agr.	Cohen κ	Weighted κ
Evaluation	90.6%	0.91	0.80
Self-regulation	86.4%	0.86	0.91
Analysis	81.7%	0.82	0.84
Inference	78.2%	0.78	0.83
Interpretation	67.3%	0.67	0.76
Explanation	65.9%	0.66	0.74
Overall	88.1%	0.78	0.83

Note. κ = Cohen's kappa. Weighted κ uses quadratic weights. N = 252 paired cells.

In the second phase, the data were analyzed according to Braun and Clarke's (2006) six-step thematic analysis procedure. First, all transcripts were read several times to familiarize ourselves with the data, and then initial codes were generated. Two types of codes were used in the coding process: deductive codes derived from Facione's six dimensions and inductive codes that captured patterns emerging directly from the data. Relationships between the codes were examined to identify potential themes; each theme was reviewed to determine whether it represented the data set as a whole; and finally, the themes were defined and reported. Each theme was supported by verbatim quotes from at least three different participants.

The analysis's trustworthiness was addressed based on Lincoln and Guba's (1985) four criteria (Nowell et al., 2017). Credibility was supported by participant validation, pilot coding, the use of two independent coders, and the resolution of disagreements by an independent referee. For transferability, the participants and context were described in detail. Dependability was ensured through a detailed coding guideline in

which all coding decisions were documented. Confirmability was strengthened by recording the researcher's log and the rationale for any revised scores.

Findings

This section first presents the participants' general critical thinking profiles. The findings are then discussed in relation to Facione's dimensions and scenarios, and thematic patterns are presented alongside participant quotes. The findings follow the order of the research questions.

Participants' Critical Thinking Profiles

Under the first research question, a scaled critical thinking score was obtained by aggregating the scores each participant received across three scenarios and six dimensions. Table 3 presents the participants' scaled scores and profiles.

Table 3. *Consensus scores and critical-thinking profiles by participant (N = 14)*

Code	Raw total	Valid n	Scaled /54	Profile
K1	38	18	38.0	Developing
K2	47	17	49.8	Deep
K3	36	18	36.0	Developing
K4	38	17	40.2	Developing
K5	43	18	43.0	Deep
K6	52	18	52.0	Deep
K7	40	18	40.0	Developing
K8	35	18	35.0	Developing
K9	51	18	51.0	Deep
K10	47	18	47.0	Deep
K11	44	18	44.0	Deep
K12	47	18	47.0	Deep
K13	52	18	52.0	Deep
K14	43	18	43.0	Deep

Note. Profile bands: 18–29 Superficial, 30–41 Developing, 42–54 Deep.

Scaled scores ranged from 35 to 52, with nine participants in the deep thinker profile, five in the developing thinker profile, and none in the superficial thinker profile (Figure 1).

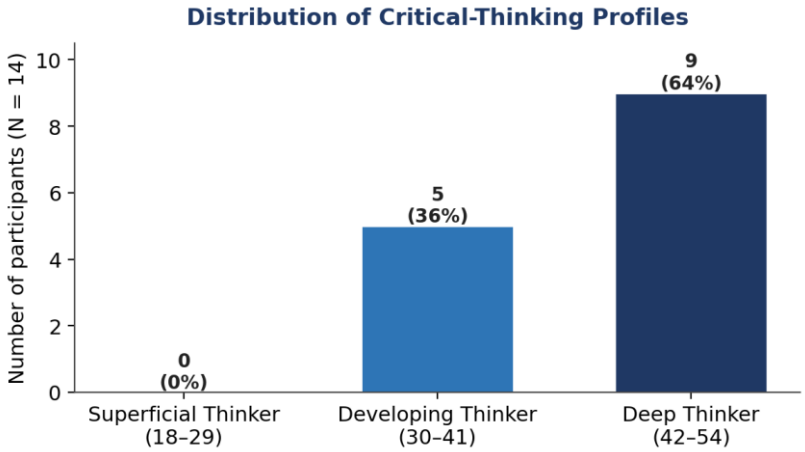

Figure 1. *Distribution of critical-thinking profiles (N = 14)*

The group generally performed well. That two-thirds of the participants fell into the deep thinker category suggests a sample relatively strong in critical reasoning, while the absence of superficial thinkers indicates that the scenarios triggered at least moderate reasoning in all students. This distribution suggests that associate-degree students with internship experience can reason at a higher-than-expected level when faced with familiar professional situations, reflecting not only individual competence but also the contextual fit of the measurement approach.

Findings by Facione Dimensions

Within the scope of the second research question, critical thinking scores were examined separately across the six Facione dimensions. Table 4 presents the scenario-based means, overall mean, and standard deviation for each dimension.

Table 4. *Descriptive statistics of consensus scores by Facione dimension (1–3 scale)*

Dimension	S1 M	S2 M	S3 M	Overall M	SD
Analysis	2.50	2.86	2.46	2.61	0.49
Self-regulation	2.07	2.71	2.92	2.56	0.55
Evaluation	2.50	2.36	2.79	2.55	0.55
Inference	2.29	2.64	2.43	2.45	0.55

Dimension	S1 M	S2 M	S3 M	Overall M	SD
Explanation	2.29	2.43	2.57	2.43	0.63
Interpretation	2.00	2.14	2.21	2.12	0.71

Note. M = mean. Items scored 0 (not probed) were excluded from the means. Overall study mean = 2.45 (SD = 0.60).

The overall mean was 2.45, with the highest mean in the Analysis dimension (2.61) and the lowest in the Interpretation dimension (2.12), as shown in order in Figure 2.

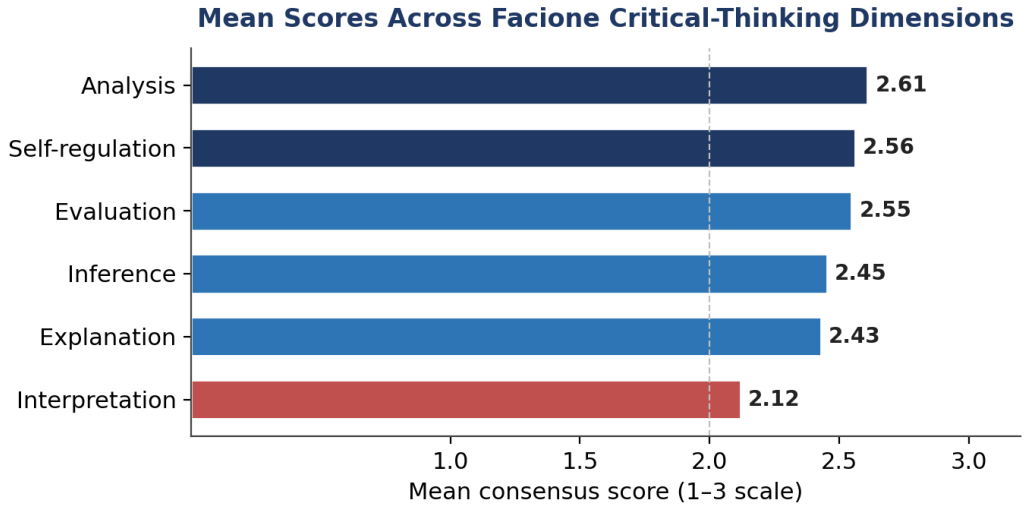

Figure 2. *Mean consensus scores across the six Facione dimensions*

Figure 2 shows that the Analysis, Self-regulation, and Evaluation dimensions cluster closely with high scores, while Interpretation lags well behind the other five. This suggests that students are relatively strong at solving problems and identifying their causes but struggle to define and frame the problem at the outset, as if the solution phase is emphasized in teaching while problem definition is neglected. Whether the low Interpretation performance reflects a teaching deficiency or a developmental stage warrants further investigation.

The distribution of dimension scores among participants is detailed in the heat map in Figure 3.

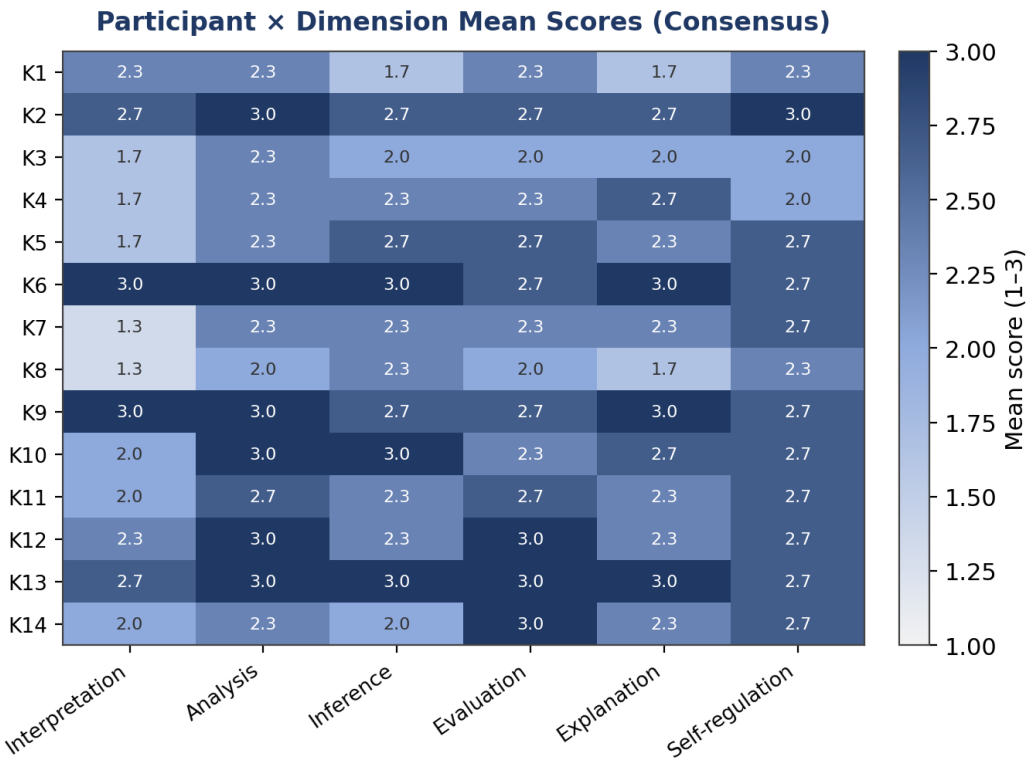

Figure 3. *Participant × dimension mean scores (darker shading indicates higher scores)*

Figure 3 shows high scores concentrated in the Analysis and Self-regulation columns, while Interpretation is represented by more moderate scores among many participants. K6, K9, and K13 scored high across nearly all dimensions, whereas K3 and K8 showed a lower, more fluctuating profile. High-performing students thus maintained consistent reasoning across dimensions while lower-performing students retained strength in some and lagged in others, suggesting that critical thinking here varies by dimension rather than functioning as a single unified skill.

Findings by Scenario

Within the scope of the third research question, findings were compared across scenarios. The scenario averages were calculated as 2.27, 2.52, and 2.56, respectively. The lowest average is observed in the first scenario, which addresses software and document incompatibility. Figure 4 presents the dimension averages by scenario.

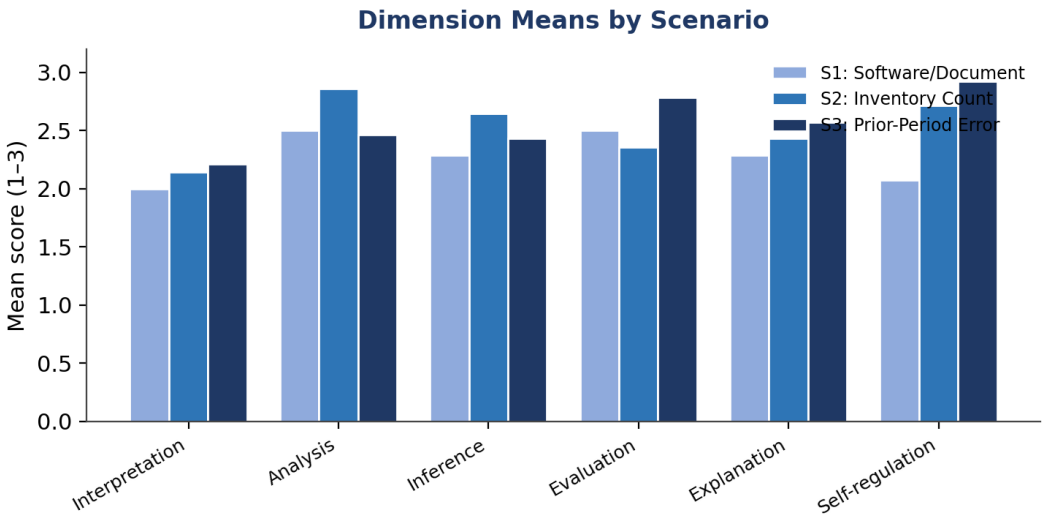

Figure 4. *Dimension means by scenario (S1: Software/Document, S2: Inventory Count, S3: Prior-Period Error)*

Figure 4 shows that the first scenario lags behind the other two, particularly in Self-regulation and Interpretation, which can be attributed to its more abstract root-cause demands. The higher Self-regulation in the third scenario suggests that a regulatory-based error triggered more pronounced self-reflection. The structure and concreteness of a scenario therefore directly influence the reasoning dimension students display, underscoring that the content of the assessment tool shapes the findings and that scenario design requires care.

Thematic Findings

As part of the fourth research question, a thematic analysis of the interview data identified five themes. Table 5 presents the themes, their corresponding codes, and the distribution of participants. The themes are explained below along with the participants' direct statements.

Evidence-Based Accounting Reasoning

The first theme was observed in all participants. Students tended to verify their claims with documentation while solving a problem. A participant investigating an inventory discrepancy systematically described their search for evidence by examining different types of records.

Table 5. *Themes derived from thematic analysis and their distribution across participants*

Theme	Associated codes	N (of 14)	Coverage
Evidence-based accounting reasoning	TEK-bilgi, SIS-düşünme	14	100%
Perceiving small discrepancies as cumulative risk	SIS-düşünme, GEL-önlem	14	100%
Self-correction through expert support	OTO-başvuru, OZD-spe	12	86%
Internal control and prevention of recurrence	GEL-önlem, SIS-düşünme	14	100%
Professional responsibility and ethical sensitivity	MES-sorumluluk, ISB-çekinme	10	71%

"First, I speak with the staff responsible for this task and the security supervisor. I would examine in detail the records of how much product arrived, how much was sold, and the inventory transactions. If the records were entered correctly and the numbers still don't match, I would review the security camera footage from that period." (K4, S2)

Similarly, another participant noted that the discrepancy could not be attributed to a single cause and evaluated potential sources at the record level.

"300 units is definitely not a small discrepancy. It could stem from a single sale or result from multiple causes. Or the product might be defective, set aside, but not deducted from inventory." (K3, S2)

From these statements, it can be said that students tend to rely on documents and records before reaching a conclusion.

Viewing Small Discrepancies as Cumulative Risks

The second theme emerged among all participants. The students realized that while discrepancies at the cent level may seem insignificant in the short term, they can accumulate over time and turn into a major risk. One participant explicitly raised this point in response to the manager's view.

"My manager might say these discrepancies are insignificant to speed up the work. However, even if this doesn't cause a problem right now, it could accumulate over time and lead to larger discrepancies. We need to take precautions now." (K2, S1)

Another participant emphasized that the cumulative effect would grow if the discrepancy recurred.

"If there are only small cent-level discrepancies on a few invoices, this might not be considered very significant. However, if a large number of invoices are processed consecutively and these discrepancies persist, the total could reach significant amounts." (K3, S1)

This indicates that students perceive not a single error but the systematic recurrence of errors as a risk.

Self-Correction with Expert Support

The third theme was observed in twelve out of fourteen participants. After acknowledging the error, students tended to consult an experienced colleague or a financial advisor. One participant expressed this approach alongside an openness to learning.

"The moment I realized my mistake, I would first consult experienced individuals well-versed in the subject, then conduct my own research. After learning the correct method, I would update my processes accordingly. By maintaining an open attitude toward learning, I would correct myself." (K4, S1)

Another participant combined the consultation process with a continuous monitoring mechanism.

"I proceed with a continuous monitoring mechanism. I also update myself by seeking support from more experienced individuals, financial advisors, or my unit supervisor." (K5, S1)

These examples suggest that students view errors not as personal failures but as learning opportunities.

Internal Control and Prevention of Repetition

The fourth theme was observed in all participants. Students suggested mechanisms such as cross-checking, a second pair of eyes, and internal audits to prevent the recurrence of errors. One participant explained their preventive approach through record-keeping discipline.

"I track inventory by categorizing it in greater detail. I remind warehouse staff to be more careful and ensure that every transaction is recorded in real time. When every transaction is documented, we can identify the source of a problem much more quickly if one arises." (K2, S2)

Another participant highlighted the necessity of a comprehensive retrospective review following the discovery of an error.

"I assess the likelihood that the former employee may have made errors in other areas as well. I thoroughly review the submitted invoices, declarations, and other financial documents from that period to check for any other inconsistencies." (K7, S3)

These statements indicate that the students are not content with a single correction but are seeking a systematic preventive approach.

Professional Responsibility and Ethical Sensitivity

The fifth theme was observed in five participants. The students demonstrated a tendency to prioritize accurate record-keeping and honest reporting despite pressure from their supervisor or manager. One participant clearly diverged from the supervisor's perspective.

"Although it may seem insignificant in the short term, I believe it can make a big difference in the long run. Unlike my supervisor, I believe this is definitely a situation that requires attention." (K1, S1)

However, some participants showed hesitation about objecting to a supervisor. One student expressed this tension based on their internship experience.

"Actually, I experienced exactly the same situation during my internship. I, too, tried to correct these discrepancies thinking that justice would prevail, but the authorities at the place where I interned did not show much interest in this matter." (K4, S1)

This suggests that students experience a tension between accurate documentation and their relationship with authority. When evaluated as a whole, the thematic findings reveal that students frequently employ evidence-seeking and systems thinking, while they weigh their sense of responsibility within the context of their relationship with authority.

Discussion

This study examined the critical thinking and problem-solving skills of associate-degree accounting students through realistic professional scenarios. In this section, the findings are discussed in relation to the literature, following the order of the research questions.

Critical Thinking Profiles

This study found nine deep thinkers, five developing thinkers, and no superficial thinkers, a level above what is typically expected for associate-degree students. This does not align at first glance with Demir et al. (2023), who found that Turkish students scored lower than peers from other countries on a standard test. The discrepancy likely reflects the measurement approach. Demir et al. (2023) used a context-free multiple-choice test, whereas this study used scenario-based assessment built on situations familiar to the students, nearly all of whom had internship experience. Realistic, field-specific scenarios therefore appear to make students' reasoning more apparent than context-free tests.

Patterns Across Dimensions

The most prominent finding is the gap between the Analysis dimension, which had the highest average, and the Interpretation dimension, which had the lowest. Students were relatively strong at breaking a problem into its components but struggled to define and frame it at the outset. This is consistent with the expert-novice research tradition. Chi et al. (1981) showed that novices represent problems through surface features whereas experts frame them through deep structure, and Schoenfeld and Herrmann (1982) found that perception shifts toward solution principles as expertise grows. The participants reasoned functionally at the solution stage but showed a novice-like pattern at the framing stage, consistent with Jones and Davidson's (2007) point that problem-solving differs in unstructured situations. From a vocational education perspective, the weakness lies not in tackling a problem but in framing it, the stage addressed least explicitly in technically oriented curricula. Problem definition should therefore be treated as a conscious instructional objective rather than a skill students acquire on their own.

A second pattern emerged across the three scenarios. The lowest average occurred in the first scenario, which required root-cause analysis of a software-document discrepancy, while the third scenario, involving a regulatory violation, yielded the highest self-regulation scores. A scenario's structure thus influences which dimension of reasoning it brings to the fore. The first scenario was the most abstract, asking students to trace an invisible systematic cause, which appears to have lowered both interpretation and self-monitoring; the rule-based third scenario more readily directed students to reflect on their own reasoning. This is consistent with the scenario-design literature, which holds that contextual framing shapes the reasoning elicited (Poulou, 2001; Rungtusanatham et al., 2011). The assessment tool therefore not only records reasoning but partly conditions which aspect of it is expressed.

This pattern has educational implications. Wolcott and Sargent (2021) argue that critical thinking in accounting education should be approached developmentally, and Dickins and Reid (2023) found that students often face questions beyond their developmental level. The low interpretation scores suggest that teaching should give more weight to problem-definition and framing. Turner and Tyler (2023) observed Facione's dimensions in an accounting course but reported no comparable imbalance, so the present study adds to the literature on patterns across dimensions.

Thematic Patterns

The thematic analysis identified five themes. The first, evidence-based reasoning, appeared in all participants: before reaching a conclusion, students traced the relevant documents and records rather than relying on first impressions. This aligns with professional skepticism, which Hurtt (2010) defined as tendencies including evidence-seeking and a questioning attitude, and corresponds to the professional-judgment layer of the framework adopted here. The participants' search for supporting records, observed before any formal auditing training, can be viewed as

an early form of this disposition, suggesting it is already present in nascent form among associate-degree students.

The second and fourth themes concern viewing small differences as cumulative risk and preventing the recurrence of errors. Together they indicate that students evaluated the problem as a systematic pattern rather than an isolated incident. Participants did not treat a minor discrepancy as a one-off error to be corrected and forgotten. Instead, they asked what the discrepancy implied for the wider set of records and how a similar error could be avoided. This forward-looking framing suggests that students did not stop at detecting the error but reasoned about its causes and consequences.

The third theme, self-correction with expert support, showed that students treated an error as a learning opportunity rather than a personal failure. After acknowledging it, they tended to consult an experienced colleague or financial advisor before deciding how to proceed. This relates directly to the self-regulation dimension in Facione's framework, in which the thinker monitors and corrects their own reasoning, though here correction was sought through expert support rather than in isolation, suggesting that students situated their judgment within a professional community of practice.

The fifth theme, professional responsibility and ethical sensitivity, presents a more complex picture. Students prioritized accurate recording and honest reporting yet expressed reservations about challenging a supervisor directly. This tension is consistent with research on authority pressure among accounting students. Owusu et al. (2020) found that students tend to report improper practices but prefer internal over external mechanisms, while Elias (2008) showed that students with high professional commitment report more readily. The participants' preference for internal resolution aligns with these findings and indicates that ethical intention and willingness to confront authority do not always move together. Read alongside Omurgonulsen and Omurgonulsen's (2009) emphasis on accounting ethics in Turkey, ethical sensitivity at this stage appears genuine but constrained by the hierarchical setting students expect to enter.

This pattern can be read within a broader international literature on critical thinking in vocational education, where developing higher-order skills in practically oriented programs is a recurring challenge. López et al. (2023) found that in Chilean technical-vocational education critical thinking developed poorly when practical-skill acquisition was prioritized, but improved when higher-education methods were adapted to the setting. Vignjević Korotaj and Buchberger (2026) likewise noted that vocational programs lack concrete frameworks for developing critical thinking despite growing labor-market demand. The framing weakness identified here fits this picture, since problem definition is the reasoning stage addressed least explicitly in technically oriented curricula. The gap among associate-degree accounting students

therefore appears to reflect a structural feature of vocational education rather than a single program or country.

The findings also speak to how reasoning is assessed in such programs. Villarroel et al. (2024) reported that higher vocational institutions rely heavily on closed-ended recall items while authentic tasks remain rare. The present results support the view that context-free tools underestimate what vocational students can do, since the scenario-based approach made visible reasoning a standardized test would likely miss. More broadly, Van Damme et al. (2023), in a performance-based study across six countries, showed that critical thinking and problem-solving can be measured reliably across cultures even as learning outcomes vary by system, supporting both the cross-cultural validity of this approach and the expectation that the profile would differ elsewhere.

Implications for Professional Education

The findings carry implications beyond accounting. At the theoretical level, the three-tiered framework adopted here, namely Facione's dimensions, the expert-novice distinction in problem representation, and the concept of professional judgment, distinguished strong from weak judgment in an authentic task with associate-degree students. This supports scenario-based assessment embedded within the field as a way to make professional judgment visible, an approach that context-free tools cannot provide and that research on authentic assessment calls for (Villarroel et al., 2024). Because the structure of the research is not specific to accounting, the framing weakness observed here is likely to recur wherever professional programs prioritize technical scope over the conscious development of judgment.

At the curriculum and policy levels, a consistent weakness in problem framing points to a specific instructional goal. Curricula that move students quickly from a given problem to its technical solution may leave the framing stage underdeveloped even when analytical processes are mastered. Incorporating unstructured scenarios and requiring students to define and reframe a problem before solving it could help, since critical thinking improves in professional settings when reasoning is explicitly addressed (López et al., 2023). Because associate-degree graduates enter the workforce with little delay, the capacity to frame unfamiliar situations is part of what makes vocational education a credible pathway to a profession (Romero-Díaz de la Guardia et al., 2024). Assessment regimes that rely on standardized, context-free testing may underestimate this capacity (Rothinam et al., 2025; Vignjević Korotaj & Buchberger, 2026).

Limitations and Future Research

Several boundaries should be kept in mind when reading these findings. The data come from fourteen students at a single vocational college, so the patterns reported here describe this group rather than associate-degree accounting students as a whole. This is a feature of the qualitative design rather than a flaw in it; the aim was an in-

depth account of how these students reason, not a statistical generalization to a wider population. A second boundary concerns the scenarios themselves. They were built around three professional situations, and accounting practice extends well beyond these. Reasoning that the scenarios did not call for, such as judgments about consolidation or tax treatment, therefore remained outside view.

This study is an attempt to reveal the judgment skills of associate-degree accounting students through realistic scenarios, and its limitations point to specific next steps. The clearest is the interpretation dimension, where participants performed least well. Whether this weakness is a stable characteristic of the group or one that responds to a short instructional intervention is a question the present design cannot answer, and a pre-post study could. Working with larger samples drawn from colleges in different regions would also show how far the profiles found here recur elsewhere. Finally, pairing the scenario-based approach with an established quantitative instrument would allow the two kinds of evidence to be read against each other, and it is often the case that convergence between them lends a finding more weight than either could carry alone.

Conclusion

This study examined the critical thinking and problem-solving skills of associate-degree accounting students through realistic professional scenarios. The findings revealed that students were relatively strong in solving problems and identifying their causes, but struggled to correctly define and frame the problem from the outset. Thematic findings, meanwhile, showed that students frequently used evidence-seeking and systems thinking, while weighing their sense of responsibility within the context of authority relationships. This pattern suggests that critical thinking operates among these students not as a single, integrated skill, but as a structure that varies across dimensions; the most prominent area for development is problem definition and framing.

This study is an effort to assess the reasoning skills of associate-degree accounting students through realistic scenarios. It can be argued that a scenario-based approach embedded within the field makes reasoning visible—something that context-free tests are likely to overlook. For these students, who enter the workforce immediately after graduation, early exposure to realistic professional situations is expected to foster critical thinking.

Ethical Approval

The administrative permissions required for this study were obtained from the relevant vocational school administration and the Social Sciences Human Research Ethics Committee of Bolu Abant İzzet Baysal University (protocol no. 2026/321). All participants were 18 years of age or older; each student signed a written consent form before the study began, and participation was voluntary. There was no relationship

between the researcher and the students regarding any course or grades. Participants' identifying information has been coded to ensure anonymity.

Conflict of Interest

The author declares that there is no conflict of interest regarding this study.

Funding

This research has not received any special support from any public, commercial, or non-profit organization.

Data Availability

Interview data supporting the findings of this study may be obtained from the corresponding author upon reasonable request, provided that participant confidentiality is maintained. Raw interview transcripts are not publicly shared due to participant confidentiality.

References

[1] Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. University of Chicago Press.
[2] Asonitou, S., & Hassall, T. (2019). Which skills and competences to develop in accountants in a country in crisis? *The International Journal of Management Education, 17*(3), 100308. https://doi.org/10.1016/j.ijme.2019.100308
[3] Baril, C. P., Cunningham, B. M., Fordham, D. R., Gardner, R. L., & Wolcott, S. K. (1998). Critical thinking in the public accounting profession: Aptitudes and attitudes. *Journal of Accounting Education, 16*(3–4), 381–406. https://doi.org/10.1016/S0748-5751(98)00023-2
[4] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
[5] Cascio, M. A., Lee, E., Vaudrin, N., & Freedman, D. A. (2019). A team-based approach to open coding: Considerations for creating intercoder consensus. *Field Methods, 31*(2), 116–130. https://doi.org/10.1177/1525822X19838237
[6] Chabrak, N., & Craig, R. (2013). Student imaginings, cognitive dissonance and critical thinking. *Critical Perspectives on Accounting, 24*(2), 91-104. https://doi.org/10.1016/j.cpa.2011.07.008
[7] Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2
[8] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

[9] Demir, C., French, B. F., & Hand, B. (2023). Cross-cultural critical thinking profiles: A multigroup latent profile analysis. *Thinking Skills and Creativity, 48*, 101286. https://doi.org/10.1016/j.tsc.2023.101286
[10] Dickins, D., & Reid, J. (2023). Integrating a foundation for the development of critical thinking skills into an introductory accounting class. *Accounting Education, 32*(3), 278–299. https://doi.org/10.1080/09639284.2022.2063025
[11] Elias, R. Z. (2008). Auditing students' professional commitment and anticipatory socialization and their relationship to whistleblowing. *Managerial Auditing Journal, 23*(3), 283–294. https://doi.org/10.1108/02686900810857721
[12] Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44–48.
[13] Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi report)*. California Academic Press.
[14] Gyekye, K. A., & Amo, O. (2024). Quality of accounting graduates and employers' expectations in Ghana. *Journal of International Education in Business, 17*(3), 395–415. https://doi.org/10.1108/JIEB-07-2023-0044
[15] Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences, 112*(8), 2395–2400. https://doi.org/10.1073/pnas.1416587112
[16] Hawk, H. (2025). Unlocking the potential: enhancing higher-order thinking skills in accounting education. *Accounting Education*, 34(6), 998-1025. https://doi.org/10.1080/09639284.2024.2375600
[17] Hurtt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory, 29*(1), 149–171. https://doi.org/10.2308/aud.2010.29.1.149
[18] Jackling, B., & De Lange, P. (2009). Do accounting graduates' skills meet the expectations of employers? A matter of convergence or divergence. *Accounting Education, 18*(4–5), 369–385. https://doi.org/10.1080/09639280902719341
[19] Jones, S. H., & Davidson, R. A. (2007). Measuring the problem-solving abilities of accounting and other business students: A comparison and evaluation of three methods. *Accounting Education, 16*(1), 65–79. https://doi.org/10.1080/09639280600826034
[20] Kimmel, P. D. (1995). A framework for incorporating critical thinking into accounting education. *Journal of Accounting Education, 13*(3), 299–318. https://doi.org/10.1016/0748-5751(95)00012-B
[21] Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications.
[22] López, F., Contreras, M., Nussbaum, M., Paredes, R., Gelerstein, D., Álvares, D., & Chiuminatto, P. (2023). Developing critical thinking in technical and

vocational education and training. *Education Sciences, 13*(6), 590. https://doi.org/10.3390/educsci13060590

[23] Nie, Y., & Mastor, N. H. (2024). Accounting employability: A systematic review of skills, challenges, and initiatives. *Cogent Business & Management, 11*(1), 2433161. https://doi.org/10.1080/23311975.2024.2433161

[24] Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1–13. https://doi.org/10.1177/1609406917733847

[25] Omurgonulsen, M., & Omurgonulsen, U. (2009). Critical thinking about creative accounting in the face of a recent scandal in the Turkish banking sector. Critical Perspectives on Accounting, 20(5), 651-673. https://doi.org/10.1016/j.cpa.2007.12.006

[26] Owusu, G. M. Y., Bekoe, R. A., Anokye, F. K., & Okoe, F. O. (2020). Whistleblowing intentions of accounting students: An application of the theory of planned behaviour. *Journal of Financial Crime, 27*(2), 477–492. https://doi.org/10.1108/JFC-01-2019-0007

[27] Papageorgiou, E. (2023). The inclusion of critical thinking in an accounting curriculum: Students' perceptions. *Journal of Education (South Africa)*, (91), 132–154. https://doi.org/10.17159/2520-9868/i91a08

[28] Paul, R. & Elder, L. (2020). *Critical thinking: Tools for taking charge of your learning and your life* (3rd ed.). Bloomsbury Publishing

[29] Payan-Carreira, R., Sacau-Fontenla, A., Rebelo, H., Sebastião, L., & Pnevmatikos, D. (2022). Development and validation of a Critical Thinking Assessment-Scale Short Form. *Education Sciences, 12*(12), 938. https://doi.org/10.3390/educsci12120938

[30] Poulou, M. (2001). The role of vignettes in the research of emotional and behavioural difficulties. *Emotional and Behavioural Difficulties, 6*(1), 50–62. https://doi.org/10.1177/1363275201006001005

[31] Romero-Díaz de la Guardia, J. J., Olmos-Gómez, M. del C., & García-Garnica, M. (2024). Skills for a working future: A model based on critical thinking in Vocational Training and Baccalaureate students. *Aula Abierta, 53*(4), 369–379.

[32] Rothinam, N., Vengrasalam, R., Naidu, S., Nachiappan, S., & Jabamoney, S. (2025). Systematic literature review on critical thinking in higher education. *Edelweiss Applied Science and Technology, 9*(5), 2046–2063. https://doi.org/10.55214/25768484.v9i5.7377

[33] Rungtusanatham, M., Wallin, C., & Eckerd, S. (2011). The vignette in a scenario-based role-playing experiment. *Journal of Supply Chain Management, 47*(3), 9–16. https://doi.org/10.1111/j.1745-493X.2011.03232.x

[34] Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 8(5), 484–494. https://doi.org/10.1037/0278-7393.8.5.484

[35] Sinnewe, E., Yao, D. (T.), & De Zwaan, L. (2023). Developing critical thinking: An examination of contemporary practices in accounting. *Accounting & Finance, 63*(1), 403–425. https://doi.org/10.1111/acfi.13054

[36] Terblanche, E. A. J., & De Clercq, B. (2021). A critical thinking competency framework for accounting students. *Accounting Education, 30*(4), 325–354. https://doi.org/10.1080/09639284.2021.1913614

[37] Terblanche, E. A. J., Shuttleworth, C. C., van Rooyen, A. A., & Masela, R. N. (2023). Critical thinking: Stakeholder expectations and challenges for accountancy educators. *South African Journal of Accounting Research, 37*(3), 225–244. https://doi.org/10.1080/10291954.2022.2148925

[38] Turner, M., & Tyler, M. (2023). Demonstrating critical thinking in accounting: Applying a competency framework. *Accounting Education, 32*(6), 713–734. https://doi.org/10.1080/09639284.2022.2105653

[39] Van Damme, D., Zahner, D., Cortellini, O., Dawber, T., & Rotholz, K. (2023). Assessing and developing critical-thinking skills in higher education. *European Journal of Education, 58*(3), 369–386. https://doi.org/10.1111/ejed.12563

[40] Vignjević Korotaj, B., & Buchberger, I. (2026). Teaching for critical thinking in vocational education and training. *International Journal of Training and Development, 30*(1), 93–103. https://doi.org/10.1111/ijtd.12378

[41] Villarroel, V., Melipillán, R., Santana, J., & Aguirre, D. (2024). How authentic are assessments in vocational education? An analysis from Chilean teachers, students, and examinations. *Frontiers in Education, 9*, Article 1308688. https://doi.org/10.3389/feduc.2024.1308688

[42] Wolcott, S. K., & Sargent, M. J. (2021). Critical thinking in accounting education: Status and call to action. *Journal of Accounting Education, 56*, 100731. https://doi.org/10.1016/j.jaccedu.2021.100731

Appendix A. Scenario-Based Interview Form

The interview form consisted of three scenarios representing authentic professional situations that associate-degree accounting graduates may encounter early in their practice. Each scenario was followed by six probe questions, each corresponding to one dimension of Facione's (1990) critical thinking taxonomy. The probes were used flexibly: when a participant had already addressed a dimension, the corresponding probe was shortened or omitted, and a minimum of four probes was administered for each scenario. The interviews were conducted in Turkish; the scenarios and probes are presented below in English translation.

Scenario 1: Software–Document Discrepancy

You transfer invoices received from the e-invoice system into the accounting software (e.g., Logo or Zirve). Continuous rounding differences of 3 to 5 kuruş, and sometimes of a few liras, arise between the VAT amount automatically calculated by the system and the VAT amount on the invoice sent by the supplier. Your accounting supervisor says, "These are negligible kuruş differences; don't waste time, just enter them and move on."

Probe 1 (Interpretation). In this situation, what do you think the real problem is? Can you define the situation in your own words?

Probe 2 (Analysis). What might the possible technical causes of this difference be? Can you list all the possibilities that come to mind?

Probe 3 (Inference). What steps would you take, and in what order, to test these causes?

Probe 4 (Evaluation). Your supervisor says that "kuruş differences are negligible." Do you agree with this view? In your opinion, what are the short- and long-term effects of these differences?

Probe 5 (Explanation). Suppose you have carried out your research and found the cause. How would you report this to your supervisor? What kind of explanation would you use?

Probe 6 (Self-regulation). If you realized at the end of your research that your initial assumption was wrong, what would you do? How would you revise your approach?

Scenario 2: Year-End Inventory Count Variance

The year-end has arrived, and a physical inventory count was carried out in the warehouse. According to the accounting records (trial balance), there should be 1,000 units of Product A in the warehouse, but only 700 units were found during the count. 300 units are missing. Your manager has asked you to investigate the cause of this difference and to prepare a report.

Probe 1 (Interpretation). What do you think the 300-unit difference means? How would you frame this situation?

Probe 2 (Analysis). What might the underlying causes of this difference be? Can you list as many possibilities as you can?

Probe 3 (Inference). Which documents or records would you examine in order to confirm or eliminate each possibility?

Probe 4 (Evaluation). If your investigation reveals more than one cause, how would you decide which one takes priority?

Probe 5 (Explanation). In what report format would you present your findings to management?

Probe 6 (Self-regulation). Is there anything you might have overlooked in this process? How would you question your own research process?

Scenario 3: Detection and Resolution of a Prior-Period Accounting Error

You have started working as an accounting clerk at a new company. While reviewing the records of previous months, you notice that the previous accountant deducted the entire VAT on the company's leased passenger vehicles, violating the VAT-deduction restriction for passenger vehicles. Your manager has asked you to correct the situation and to prepare an action plan.

Probe 1 (Interpretation). When you detect this error, what is the first thing you try to understand?

Probe 2 (Analysis). Which records and which periods would you examine in order to determine the scope of the error?

Probe 3 (Inference). How would you calculate the financial impact of the error (additional tax, late-payment interest, and penalties)?

Probe 4 (Evaluation). How would you weigh the advantages and disadvantages of filing an amended tax return?

Probe 5 (Explanation). How would you present the action plan you have prepared to management?

Probe 6 (Self-regulation). Would you consider that there might be other overlooked errors in this process? How would you check for this?

Appendix B. Scoring Rubric for the Six Critical Thinking Dimensions

Each probe was mapped to a single dimension of Facione's (1990) framework and scored on a three-point scale: 1 (low), 2 (medium), and 3 (high). Off-topic or absent responses were scored 0. At boundary cases, the indicator criteria in the right-hand column were counted, and the level meeting the greater number of criteria was assigned.

Dimension 1: Interpretation. *How the participant grasps and frames the problem.*

Score	Description	Indicators
1 (Low)	Defines the problem superficially and one-dimensionally.	Restates the scenario; does not question its scope or context; uses no technical terms or uses them incorrectly.
2 (Medium)	Grasps the problem partially; is aware of more than one dimension.	Notes that the problem may have several causes; establishes the context partially but cannot fully frame it; generally uses technical terms correctly.
3 (High)	Grasps the problem multidimensionally, contextually, and systemically.	Distinguishes whether the issue is structural or incidental; considers the perspectives of different stakeholders; frames the problem across regulatory, technical, and process layers.

Dimension 2: Analysis. *Capacity to break the problem into components and generate multiple causes.*

Score	Description	Indicators
1 (Low)	States a single cause or none.	Offers one possibility only; does not classify causes; establishes no relations among possibilities.
2 (Medium)	Lists two or three causes but does not link or prioritize them.	Names several possibilities in no particular order; defines some causes broadly and others narrowly; attempts categorization but incompletely.
3 (High)	Categorizes, relates, and prioritizes causes.	Considers causes across layers (technical, human, procedural); gives examples within each category; weighs or ranks the probabilities.

Dimension 3: Inference. *Ability to convert hypotheses into a systematic action plan.*

Score	Description	Indicators
1 (Low)	Proposes a vague, unworkable, or single-step plan.	Uses general statements such as "I would look into it"; presents no logical order among steps; cannot specify which source to consult.
2 (Medium)	Lists several reasonable steps but with weak sequencing logic.	Identifies the correct sources but does not order them; does not prioritize from simple to complex; links steps only partially.
3 (High)	Constructs a systematic, ordered, and prioritized action plan.	Tests the simplest possibility first (the least costly path); anticipates the likely outcome of each step; specifies a contingency plan for alternative scenarios.

Dimension 4: Evaluation. *Capacity to question arguments, produce counter-arguments, and compare alternatives.*

Score	Description	Indicators
1 (Low)	Accepts a view without questioning it; reasons one-sidedly.	Uses affirming statements; produces no counter-argument; shows no alternative perspective.
2 (Medium)	Produces arguments for and against but cannot make them concrete.	Offers general statements (e.g., short-term versus long-term); lists advantages and disadvantages without concrete examples; does not bring the discussion to a conclusion.
3 (High)	Discusses arguments with concrete evidence and reaches a deliberate decision.	Uses concrete data or regulatory references; distinguishes short- and long-term effects; concludes the discussion with a justified decision.

Dimension 5: Explanation. *Ability to present findings in a structured and persuasive manner.*

Score	Description	Indicators
1 (Low)	Cannot state the conclusion or leaves it vague.	Uses general statements such as "I would explain it"; gives no thought to presentation format; does not consider the audience.
2 (Medium)	Explains in broad terms but offers no structured format.	Says "I would write a report" but cannot detail its content; refers to some supporting data; considers the audience only partially.
3 (High)	Draws a structured, persuasive, and audience-appropriate presentation plan.	Offers a concrete report structure (summary, findings, recommendations); uses visual support such as tables or charts; selects language suited to the audience's level of knowledge.

Dimension 6: Self-regulation. *Capacity to question one's own reasoning, notice possible errors, and correct them.*

Score	Description	Indicators
1 (Low)	Shows no self-criticism; assumes everything was done correctly.	States that nothing further is needed; does not question own assumptions; feels no need for external input.
2 (Medium)	Admits a possible error but cannot specify it.	Uses a general statement such as "I may have missed something"; cannot identify blind spots; the revision strategy remains unclear.
3 (High)	Identifies specific blind spots and offers a concrete correction strategy.	States which assumption could be mistaken; selects a different starting point or tries a different method; suggests seeking external input (e.g., a certified public accountant or a colleague).

Scoring rules. Responses were scored on content rather than length. When a participant expanded an answer after a probe, the expanded answer was scored. When a participant demonstrated the same dimension across different questions, the corresponding question was scored. At boundary cases, the level meeting the greater number of indicator criteria was assigned. Off-topic or absent responses were scored 0.