

The Relationship Between Integrated Assessments and Critical Thinking Skills of First-Year Accounting Students

Melanie Cloete

Durban University of Technology
Durban, South Africa

Kansilembo Aliamutu

University of South Africa
South Africa

Akinola Aruwaji

Durban University of Technology
Durban, South Africa

Abstract

In the intelligent era, characterized by rapid technological advancements, the accounting profession is shifting from routine tasks to demanding higher-order cognitive skills, particularly critical thinking. This study examines the relationship between integrated assessments and the development of critical thinking skills in first-year accounting students. Using a quasi-experimental design, the study examines the impact of integrated assessments on the critical thinking abilities of two groups of accounting students: an experimental group exposed to integrated assessments and a control group receiving traditional assessments. The Watson–Glaser Critical Thinking Appraisal, UK Edition was used to measure critical thinking skills in both groups. Using analysis of variance (ANCOVA), the findings indicate that students exposed to integrated assessments scored significantly higher on critical thinking measures than those in the control group. These findings underscore the importance of incorporating such assessments to equip accounting graduates with the necessary skills to navigate the complexities of the modern business landscape. This research contributes valuable empirical evidence supporting the use of integrated assessments as an effective pedagogical approach to foster critical thinking skills in accounting education.

Keywords: integrated assessments, critical thinking skills, industry-relevant skills, accounting students, higher education.

1. Introduction

In today's rapidly evolving technological landscape, critical thinking skills are no longer merely desirable but essential for success in the workplace (Organization for Economic Co-operation and Development, 2019). This is particularly true for professions such as accounting, where the increasing prevalence of artificial

intelligence and automation is transforming traditional job roles and demanding a new set of skills from graduates (Kroon et al., 2021; Jackson et al., 2023). While technical accounting knowledge remains important, the ability to analyze information, solve complex problems, and make sound judgments under pressure is becoming increasingly crucial for accounting professionals to thrive in this dynamic environment (Terblanche & De Clercq, 2021).

However, a significant gap exists between the demand for critical thinking skills in accounting graduates and the perceived effectiveness of higher education institutions in fostering these skills (Willingham, 2008; Sin et al., 2015; Nwosu & Vorster, 2021). Numerous studies have highlighted the shortcomings of traditional teaching and assessment methods in adequately preparing students for the critical thinking demands of the modern workplace (Barac & Du Plessis, 2014; De Klerk et al., 2025). This discrepancy highlights the need for innovative pedagogical approaches that can effectively bridge this gap and equip accounting graduates with the essential critical thinking skills required to succeed in their careers. Integrated assessment, which emphasizes the use of authentic, real-world scenarios and tasks, has emerged as a promising approach to fostering critical thinking skills in higher education (Cloete, 2018). By simulating the complexities and challenges of professional practice, integrated assessment encourages students to apply their knowledge and skills in meaningful contexts, promoting deeper learning and the development of higher-order thinking abilities.

This study aims to examine the relationship between integrated assessment and the development of critical thinking skills in first-year accounting students at a university of technology in South Africa. The findings of this study have the potential to make several significant contributions. Firstly, the findings provide empirical evidence on the effectiveness of integrated assessments in developing students' critical thinking, informing curriculum design to better prepare graduates for professional demands. Secondly, they contribute to the broader discourse on assessment practices in higher education and their role in fostering essential graduate attributes, particularly critical thinking. Finally, by highlighting the importance of critical thinking skills in the accounting profession, this study encourages further research and innovation in pedagogical approaches that can effectively develop these skills in future accounting professionals.

1.1 Defining Critical Thinking: Cognitive and Dispositional Dimensions

Critical thinking, defined by Dewey as reflective and reasoned judgment, encompasses both cognitive skills and dispositional attitudes (Dewey, 1997; Ennis, 1985; Facione, 2011; Lia, 2011). Cognitive skills include analysis, synthesis, and evaluation, while dispositions involve open-mindedness and a commitment to reason (Ku, 2009). Although both are important, this study focuses on cognitive skills due to the limitations of the Watson-Glaser Critical Thinking Appraisal (W-GCTA), which does not adequately assess dispositional aspects.

1.2. Critical Thinking and the Digital Transformation of the Accounting Profession

The accounting profession is in the midst of a digital transformation, marked by the increasing adoption of artificial intelligence (AI) tools (Kroon et al., 2021). This technological shift, driven by AI, is reshaping the role of accountants. As AI takes over traditional, routine accounting tasks, accountants are pivoting toward greater emphasis on providing business advisory services (South African Institute of Professional Accountants, 2019). This shift emphasizes the growing importance of critical thinking as an essential skill for accountants navigating this digital era (Hansson, 2019; Shaw et al., 2020). The South African Institute of Professional Accountants (2019) recognizes critical thinking as a crucial competency for “future-ready professional accountants”.

Employers increasingly seek graduates who possess strong critical thinking skills, as these skills are essential for navigating complex, real-world scenarios that arise in a technology-driven environment. Consequently, university programs must focus not only on technical knowledge but also on cultivating pervasive skills such as critical thinking to prepare students for the demands of the modern workplace (De Klerk et al., 2025). This shift is necessary to ensure that future accountants can make responsible and creative decisions in an increasingly digital landscape.

2. Literature Review

The literature review explores two main themes: learner assessment in higher education and critical thinking and its relationship with integrated assessments.

2.1 Learner Assessment in Higher Education

At higher education institutions, educators use different types of assessments to make judgments about the students’ level of achievement. Moreover, these judgments can have a tremendous impact on the students’ career paths. The type of assessment employed by educators influences students’ study habits, learning preferences, depth of understanding, knowledge acquisition, and their ability to apply theory in practice (Saifuddin et al., 2025). Assessments within the higher education context can be classified into two categories, namely, traditional/teacher-centered and learner-oriented/student-centered assessments.

Traditional assessments are an extension of the teacher-centered paradigm, where educators transmit information to passive students, who are then required to recall this information at a later stage (Amirova, 2025). Boud and Falchikov (2006) concurred that students are not seen as active agents in the assessment process, which is imposed on them by educators. Assessment tasks are often decontextualized, and emphasis is placed on correct answers (Huba & Freed, 2000). Furthermore, according to Huba and Freed (2000), assessment tasks are not interdisciplinary but rather focus on a single discipline.

The constructivist approach, which supports lifelong learning through learner-oriented teaching and assessment practices, has become an integral part of the South African higher education system. The assessment policy of the university of

technology (UoT) where the study took place is aligned with this constructivist approach. It views student-centeredness as the crucial pillar that informs all teaching, learning, and assessment practices at the university (Centre for Excellence in Learning and Teaching, DUoT, 2014). As mentioned previously, although the constructivist approach has been emphasized, teacher-centered approaches, which restrict critical thinking, still dominate teaching and assessment practices at institutions of higher learning (Lombard & Grosser, 2008). Educators prefer the traditional assessment methods due to large classes and time constraints resulting from research and other administrative requirements (Lombard, 2008; Lombard & Grosser 2008; Lombardi, 2008). In addition, educators may be reluctant to change their assessment methods, since this can be more challenging than the actual assessment activity. McPhun (2010) argued that educators cling to traditional assessment methods because they are unwilling to move out of their “comfort zones”.

Learner-oriented assessments extend the student-centered paradigm, positioning students as active participants and educators as facilitators (Mehdizade, 2025). This approach requires students to construct knowledge through communication, critical thinking, and problem-solving skills. Learner-oriented assessments are contextualized and interdisciplinary. They focus on problems encountered in the real world, requiring students to use their knowledge from various disciplines to solve the problem. The emphasis of the student-centered paradigm is on student engagement.

De Klerk et al. (2025) noted that if students are not actively involved in the learning process, their adaptability to the work environment is severely restricted. Active engagement in the learning process is necessary to develop students’ critical thinking skills. It is therefore vital that educators keep students engaged at all times. In addition, students’ assessment preferences can also have an impact on their learning. Saifuddin et al. (2025) maintained that students use surface learning when studying for written examinations, as opposed to deeper learning when compiling assessment essays. Furthermore, students prefer alternative assessment methods that result in deeper understanding, for example, assessment essays, instead of written examinations.

The UoT’s assessment policy, which is based on the philosophy of student-centeredness, also promotes the use of various methods of assessment, particularly those types of assessment that encourage critical thinking and problem-solving in “*discipline-or profession-specific contexts*” (Centre for Excellence in Learning and Teaching, DUoT, 2014, p. 2). According to Lombard (2008), learner-oriented assessments are an approach for enhancing students’ critical thinking skills. Learner-oriented assessments promote critical thinking, unlike traditional methods, by using interdisciplinary, real-world tasks. This study focuses on authentic and integrated assessments that develop higher-order thinking skills. As this study was conducted at a UoT, its assessment policy provides the framework for integrated assessment. The policy defines integrated assessment as a holistic approach that combines multiple skills, outcomes, and methods, often drawing on diverse evidence sources (Centre for

Excellence in Learning and Teaching, DUoT, 2014). It further promotes the use of integrated assessments across modules and at different levels of study.

2.2 Critical Thinking and Its Relationship with Integrated Assessments

Critical thinking skills are not inherent; they must be taught and practiced, that is, reinforced with appropriate assessments. This study will not focus on the teaching of critical thinking skills but rather on assessments and, in particular, the relationship between integrated assessments and the critical thinking skills of first-year university students. Since the Watson-Glaser Critical Thinking Appraisal, UK Edition (W-GCTA) assessment instrument is used to gauge the critical thinking skills of first-year accounting students, the definition presented by Watson and Glaser (2002) will be used for this paper. Watson and Glaser (2002) defined critical thinking as:

“(1) attitudes of enquiry that involve an ability to recognize the existence of problems and an acceptance of the general need for evidence in support of what is asserted to be true; (2) knowledge of the nature of valid inferences, abstractions and generalizations in which the weight or accuracy of different kinds of evidence is logically being determined; and (3) skills in employing and applying the above attitudes and knowledge.” (p. 71)

Critical thinking is closely linked to constructivist, learner-centered approaches that emphasize problem-based learning and lifelong learning (Bhardwaj et al., 2025). In South Africa, this is reinforced by the Critical Cross-Field Outcomes, where critical thinking is a core requirement embedded in all qualifications (South African Qualifications Authority, 2006). However, traditional, teacher-centered approaches still dominate practice and may limit critical thinking development (Lombard & Grosser, 2008). Aligned with this, Paul and Nosich (2013) argued that effective assessments should promote knowledge transfer, interdisciplinary integration, and workplace-relevant decision-making skills.

De Klerk et al. (2025) highlighted the need to renew assessment practices to meet evolving workplace demands. They emphasized that integrated, cross-disciplinary assessments enhance higher-order thinking, while stand-alone assessments tend to fragment learning and limit students' ability to see connections across knowledge areas. Vanderheide and Walkington (2008) argued that overreliance on stand-alone assessments can discourage critical thinking, as students often adopt surface learning strategies to cope with heavy workloads. In contrast, their study on a multidisciplinary integrated assessment with nursing students found that it enhanced theory-practice integration, reduced assessment burden, and promoted deeper learning. Increased collaboration among academics further enriched the learning experience, providing strong empirical support for the value of integrated assessments.

In the South African context, McPhun (2010) developed and implemented an integrated assessment model, finding it to be a practical and valuable learning experience. Graduates reported improved workplace preparedness, and employers expressed satisfaction with their job-readiness. A limitation of the national study is

that it did not measure the development of critical thinking skills in students but rather placed emphasis on student and employer perceptions of the integrated assessment. Prior studies indicate that integrated assessments promote deeper learning and higher-order thinking skills. Vanderheide and Walkington (2008) indicated that multidisciplinary tasks enhance learning connections and student engagement, while McPhun (2010) highlighted their role in improving work-readiness. More recent research supports these findings, with Connolly et al. (2023) emphasizing that integrated and authentic assessments foster sustainable learning through knowledge application, and Jackson (2019) demonstrating their contribution to employability skills, including critical thinking.

Despite these insights, there is limited empirical research that quantitatively measures the impact of integrated assessments on critical thinking, particularly within the South African accounting education context. While De Klerk et al. (2025) have started addressing this gap, evidence using standardized critical thinking instruments remains scarce. This study contributes by providing empirical evidence of the relationship between integrated assessment and critical thinking using a standardized measurement tool.

3. Research Design

A quasi-experimental pre-test/post-test non-equivalent groups design was adopted for this study to investigate the influence of integrated assessments on the critical thinking abilities of first-year accounting students at a UoT in South Africa. Participants were strategically allocated to either an experimental or control group, rather than through random assignment, a common feature of quasi-experimental approaches as outlined by Reichardt (2009). This intentional grouping reflects practical constraints typical of educational research conducted in authentic learning environments.

At the start of the study, both groups completed a pre-test assessment (W-GCTA UK Edition) aimed at gauging their baseline critical thinking skills. Subsequently, the experimental group engaged with an integrated assessment intervention, consisting of authentic, interdisciplinary tasks intended to enhance higher-order cognitive abilities. In contrast, the control group continued to follow conventional assessment practices without exposure to the integrated assessment. Upon completion of the intervention, all participants underwent a post-test assessment (W-GCTA UK Edition) measuring their critical thinking development. This research design enabled the evaluation of changes within and across groups over time, while addressing initial differences between groups by statistically controlling pre-test scores using analysis of covariance (ANCOVA). Although the absence of randomization presents some limitations regarding internal validity, the design remains well-suited for examining educational interventions in naturalistic academic contexts, offering valuable insights into the role of integrated assessments in strengthening students' critical thinking competencies.

3.1 Participants and Sampling

This study examines an integrated assessment built around a real-world problem, developing students' critical thinking skills. The assessment was introduced to first-year extended curriculum program (ECP) students in the Department of Management Accounting at a South African UoT. Its design was informed by the exit-level outcomes of the National Diploma in Cost and Management Accounting, as well as the Critical Cross-Field Outcomes (CCFOs), with the intention of supporting students as they progress in their studies.

The study involved first-year accounting students organized into an experimental and a control group. The experimental group consisted of students enrolled in the Cost and Management Accounting (ECP) program, while the control group comprised students from the Accounting (ECP) program in the Department of Financial Accounting. Only ECP students were included in the study. In line with Department of Higher Education and Training guidelines, which limit ECP enrolment to approximately 20% of total program intake, the number of participants was relatively small, with around 40 students per qualification. A census approach was adopted, as data were collected from all students registered in the selected programs (Saunders et al., 2019). While this allowed for full coverage of the group under study, the limited sample size may restrict the broader generalization of the findings. Ethical standards were upheld throughout the study. Participation was voluntary, informed consent was obtained prior to administration of the W-GCTA, and all test results were handled confidentially and anonymously.

3.2 Sampling Strategy and Group Allocation

The participants were selected from a single institution using a purposive sampling strategy to ensure that they had similar academic backgrounds and exposure to the course material. In total, 48 students took part in the study, with 24 assigned to the experimental group and 24 to the control group. A quasi-experimental design was employed, relying on pre-existing class groupings rather than random assignment. Students were therefore allocated to the experimental and control conditions based on their original class placement. To reduce potential selection bias, both groups were drawn from Management Accounting and Financial Accounting programs at the same academic level, where there was substantial overlap in first-year modules and prior content exposure.

The integrated assessment comprised five individual assignments. An overview of each assignment is presented in Figure 1

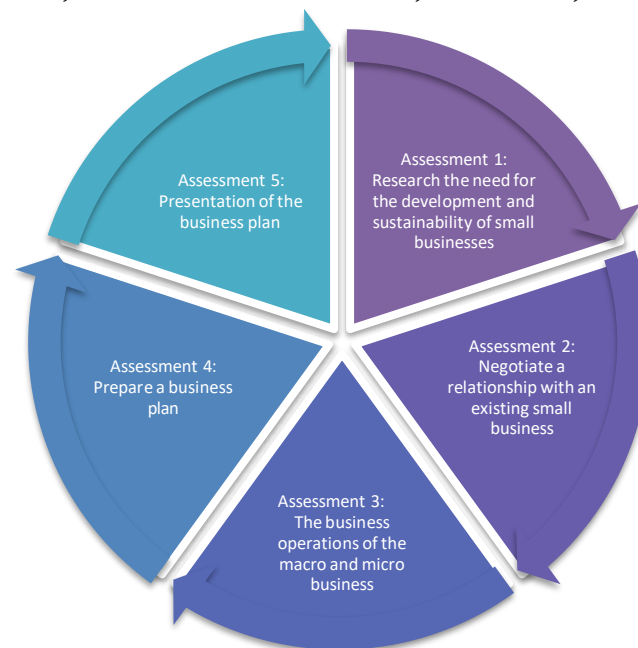


Figure 1: Overview of the integrated assessment

The integrated assessment was a *group assessment*. This ensured that students learned to work collaboratively, which is a necessary skill required in the work environment (De Klerk et al., 2025; Lombardi, 2008). The assignments had to be completed in the specified order, with each assignment providing a scaffold for the next. This scaffolding ensured that students could see the link between the individual assessments and the overall requirement of designing a business plan for a chosen small business. This general requirement (see Table 1) was aligned to the exit-level outcome for the National Diploma: Cost and Management Accounting, that is, graduates being capable of practicing as accountants and rendering accounting services to commerce, industry, and the public sector.

Table 1: Requirements of the individual assessments

Assessment number	Assessment requirements
1	Research the need for the development and sustainability of small businesses Each group was required to write a report highlighting the need for the development and sustainability of small businesses. This report had to include the following: - the most recent developments with regard to small business development and sustainability; and - at least two examples of small businesses that have had an impact on the economy and on community upliftment.
2	Negotiate a relationship with an existing small business Each group was required to prepare a report providing the details of establishing a relationship with a chosen small business. The chosen small business had to be in the surrounding area of the UoT under study. This report had to: document all the businesses approached and briefly outline the reasons why specific businesses did not want to participate;

	• describe the problems encountered and indicate how these problems were resolved, in terms of establishing a relationship with a small business; • summarize the reasons or motivation used to establish the relationship; and • provide a letter of consent, which can be written or recorded (if verbal), clearly indicating the business's willingness to participate. Photographs and videos must be provided as evidence of the established relationship.
3	The business operations of the macro and micro business The students went on an arranged excursion to a macro business within the production industry. Before the excursion, each group was given a questionnaire, which elicited information about the business operations. After the excursion, each group was required to prepare a report. This report had to: • highlight the observations made on sound business practices; • highlight observations made about the factory layout, safety measures, cleanliness, etc.; • explain the company's mission statement; and • include a basic mission statement drafted for the chosen small business.
4	Prepare a business plan Students had to prepare a business plan with a follow-up report suggesting possible ways in which the chosen small business can improve its day-to-day operations and, consequently, make the transition from survivalist to a profitable entity that will exist in the foreseeable future. The business plan had to include the following: • description of the business; • financial plan; • operational plan; and • recommendations and conclusions.
5	Presentation of the business plan The business plan was to be presented using a Microsoft PowerPoint presentation. All group members were expected to take part in the presentation. Marks were awarded for the individual as well as group presentations.

The integrated assessment required students to solve real-world problems, work collaboratively, and meet deadlines, which are key workplace skills, thereby encouraging them to think like practitioners rather than students.

3.3 Data Collection Instrument: W-GCTA

Quantitative data were collected using the W-GCTA UK Edition. The W-GCTA instrument is a multiple-choice test that consists of five subtests that measure the cognitive skills of inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. Each of the five subtests requires respondents to: make inferences by determining the correct conclusions to be drawn from the factual statements provided; identify the assumptions implied in the statements given; determine if the proposed findings follow reasonably from the given information; provide an interpretation of the information to determine if the proposed conclusions drawn are warranted; and evaluate the strength and applicability of the proposed arguments. The W-GCTA UK Edition comprises 80 questions, with 16 questions for

each subtest. The subtest questions are discipline-neutral and use scenarios based on social, political, and economic issues encountered daily (Ku, 2009). The test consists of neutral and controversial material, indicating respondents' ability to think critically, irrespective of whether they have strong feelings about the issue at hand (Watson & Glaser, 2002). Each subtest has specific instructions, and the duration of the entire test is 60 minutes.

The W-GCTA UK Edition was used for the study as it had previously been used in South Africa and several other countries and settings (Watson & Glaser, 2002). A discipline-specific critical thinking test was not available within the South African context. Despite the test not being discipline-specific, it required students to use critical thinking skills needed within the higher education context (Bernard et al., 2008; Hassan & Madhum, 2007). The five subtests enabled educators to ascertain deficiencies in students' critical thinking skills, thereby informing appropriate future assessments (Bernard et al., 2008).

This study was longitudinal as all five subtests were conducted with the control and experimental groups in March and September. The March and September timeframes were chosen as they represented pre- and post-integrated assessments for the experimental group. The aim was to measure the relationship between the integrated assessment and students' critical thinking skills. Since the experimental group was exposed to the integrated assessment in the form of an intervention, their post-test scores were expected to be significantly higher than those of the control group.

3.4 Data Collection Procedure

Data collection took place at two different stages to track changes in students' critical thinking over time. The first round of data collection was conducted in March through a pre-test administered at the start of the academic period, followed by a post-test in September after the intervention had been completed. This approach made it possible to observe how students' performance developed across the study period. Both the experimental and control groups completed the assessments under comparable conditions to ensure consistency. Ethical principles were maintained throughout, with participation being voluntary, informed consent obtained in advance, and all student test results being handled confidentially and anonymously.

3.5 Reliability and Validity

3.5.1 Reliability of the instrument

The W-GCTA is widely regarded as a reliable and valid instrument, having been applied across a range of educational and professional settings. Earlier research has reported acceptable levels of internal consistency, with Cronbach alpha values for the overall scale generally falling between 0.70 and 0.85 (Watson & Glaser, 2002). The instrument has also been used in South Africa, which supports its suitability for assessing critical thinking among students in higher education in South Africa. Relying on a well-established and standardized measure helped to ensure that the results are both consistent and dependable, thereby strengthening confidence in the findings.

3.5.2 Validity considerations

Several measures were taken to support the validity of the study. Internal validity was strengthened by working with groups that were similar at the start of the study and by applying the same standardized instrument to both groups. This approach helped ensure that any differences observed are more likely to be linked to the intervention itself rather than to inconsistencies in measurement. The study also benefited from a degree of external validity, as the intervention was carried out in a real classroom environment, making the findings more relevant to typical educational settings. Prior to the main data collection, the instrument and procedures were pilot-tested to confirm that they were clear and appropriate for the participants. Despite these efforts, certain limitations remain. The use of a quasi-experimental design, based on existing class groups, means that participants were not randomly assigned. As a result, there may be underlying differences between the groups that were not fully controlled for, and this should be considered when interpreting the findings.

4. Results

The effect of the integrated assessment intervention on students' critical thinking was examined using a combination of descriptive statistics, Pearson correlation, and ANCOVA. Descriptive measures were first used to summarize the data and determine whether the groups were comparable at the outset. ANCOVA was then employed to assess the impact of the intervention while accounting for any initial differences in critical thinking ability. In addition, correlation analysis was used to explore how the different components of critical thinking relate to one another, as well as the connection between pre-test and post-test performance. Taken together, these approaches offer a well-rounded view of both the effectiveness of the intervention and the relationships among the underlying skills.

4.1 Baseline Comparability Between Groups

An independent samples *t*-test was carried out to examine whether the experimental and control groups differed before the intervention (Table 2). The results show that the two groups had very similar pre-test scores (experimental group: $M = 52.4$, $SD = 9.8$; control group: $M = 51.9$, $SD = 9.5$). This indicates that there were no meaningful differences between the groups at the starting point.

Table 2: Descriptive statistics for test scores

Group	n	Pre-test mean (SD)	Post-test mean (SD)	Mean change
Experimental	24	52.4 (9.8)	61.2 (10.1)	+8.8
Control	24	51.9 (9.5)	53.8 (9.7)	+1.9

As reflected in Table 2, both groups began at a comparable level, but their progress differed over time. The experimental group showed a noticeable improvement in scores, whereas the control group experienced only a slight increase. This pattern suggests that the intervention may have contributed to the greater gains observed in the experimental group.

4.2 Correlation Analysis

To further explore the relationships among the various components of critical thinking, Pearson correlation analyses were conducted between the subtests and between pre- and post-test scores. Correlations were used to determine whether two or more variables were associated with one another in some way. The following correlations were done for the experimental group using Pearson correlation: a correlation between the various subtests; and an overall correlation between the pre- and post-test *t*-scores. Positive numbers indicate a positive correlation, whereas negative numbers indicate a negative correlation. A positive correlation reflects a direct relationship, while a negative correlation reflects an indirect relationship. The results of the Pearson correlation between the subtests are presented in Table 3.

Table 3: Correlations between subtest scores

		1	2	3	4	5
1. Inferences	Pearson correlation	1				
	Sig. (2-tailed)					
	N	24				
2. Recognition	Pearson Correlation	.504*	1			
	Sig. (2-tailed)	.012				
	N	24	24			
3. Deduction	Pearson correlation	.469*	.351	1		
	Sig. (2-tailed)	.021	.093			
	N	24	24	24		
4. Interpretation	Pearson correlation	.230	.293	.154	1	
	Sig. (2-tailed)	.280	.164	.472		
	N	24	24	24	24	
5. Evaluation	Pearson correlation	.572**	.352	.555**	.022	1
	Sig. (2-tailed)	.003	.092	.005	.920	
	N	24	24	24	24	24

Note: * $p < 0.05$, ** $p < 0.01$ (2-tailed)

The results indicate a strong positive relationship between evaluation of arguments (subtest 5) and both inference (subtest 1) and deduction (subtest 3), with these associations being significant at the 1% level. In addition, inference is significantly related to recognition of assumptions (subtest 2) and deduction (subtest 3) at the 5% level. These patterns suggest that certain components of critical thinking tend to

develop together, particularly those linked to reasoning and evaluative judgment. The relationship between overall pre-test and post-test *t*-scores for the experimental group is presented in Table 4. The data show a moderate positive correlation between pre-test and post-test scores ($r = .475$), which is statistically significant. This suggests that students who performed well at the outset tended to achieve higher scores after the intervention, although the relationship is not so strong as to imply that outcomes were solely determined by initial ability.

Table 4: Overall correlations between pre- and post-test *t*-scores: experimental group

		1	2
1. Pre-test <i>t</i>-score	Pearson correlation	1	.475*
	Sig. (2-tailed)		.019
	N	24	24
2. Post-test <i>t</i>-score	Pearson correlation	.475*	1
	Sig. (2-tailed)	.019	
	N	24	24

Note: * $p = 0.019$

4.3 ANCOVA Analysis

Table 5 presents the results of the ANCOVA assumption test. The test was performed to ensure the appropriateness of the analysis.

Table 5: Results of the ANCOVA assumption test

Assumption	Test statistic (<i>p</i>-value)
Normality of residuals	0.839 (Shapiro–Wilk)
Homogeneity of variance	0.806 (Levene’s test)
Homogeneity of regression slopes (interaction: pre-test × group)	0.099

The normality of residuals was assessed using the Shapiro–Wilk test. The obtained *p*-value of 0.8386 was well above the 0.05 threshold, suggesting that the residuals were normally distributed. Satisfying the normality assumption is important, as it underpins the validity of the significance tests and effect estimates produced by ANCOVA. However, the assumption of equal variances between groups was examined through Levene’s test. The test yielded a *p*-value of 0.8059, which indicates that the variance of post-test scores did not differ significantly between the experimental and control groups. Maintaining homogeneity of variance is crucial because it prevents distortion of the *F*-ratios and safeguards against biased statistical conclusions.

Moreso, the homogeneity of regression slopes was checked by investigating the interaction between pre-test scores and group membership. The *p*-value for the interaction term (pre-test × group) was 0.0987, which is greater than 0.05. Although relatively close to the threshold, it still indicates that the relationship between the covariate (pre-test score) and the dependent variable (post-test score) was consistent

across groups. This assumption is vital because it ensures that the adjustment made for initial differences applies similarly to all participants, enabling a fair comparison between groups. Finally, the results of the assumption tests confirm that all conditions required for conducting a valid ANCOVA were met. Consequently, the statistical conclusions regarding the effect of integrated assessments on students' critical thinking skills can be considered robust and reliable.

Figure 2 shows the linear association between pre- and post-test scores for both the experimental and control groups. The slopes of the regression lines are approximately parallel, supporting the assumption of homogeneity of regression slopes required for ANCOVA. The experimental group demonstrates a generally higher post-test performance compared to the control group, reflecting the positive effect of the integrated assessment intervention.

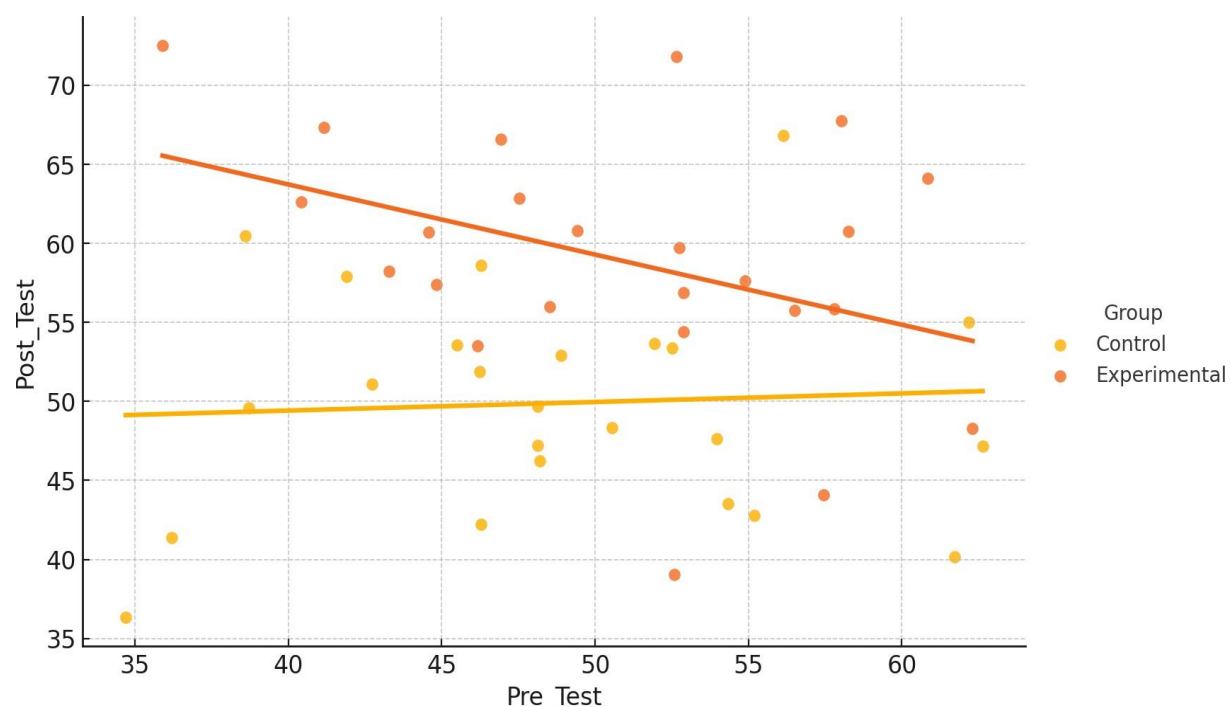


Figure 2: Linearity check on pre- and post-test scores by group

4.4 ANCOVA Results

Assessing whether the intervention leads to improvements beyond students' starting levels is key to understanding its value. Table 6 presents the ANCOVA results on the effect of the integrated assessment approach on post-test critical thinking scores, taking prior ability into account. The results show a noticeable difference between the experimental and control groups, with $F(1, 45) = 8.71$, $p = 0.005$, indicating that the intervention influenced performance even after adjusting for baseline scores. The effect size (partial $\eta^2 = 0.162$) suggests that the intervention accounts for a meaningful share of the variation in outcomes (roughly 16%).

Table 6: ANCOVA analysis

Source	Sum of squares	df	Mean square	F	p-value	Partial η^2	Effect size
Group (intervention)	801.84	1	801.84	8.71	0.005	0.162	Large
Pre-test (covariate)	1298.34	1	1298.34	14.12	<0.001	0.239	Large
Error	4310.23	45	95.78				
Corrected total	6610.41	47					

Initial performance also played a significant role, with $F(1, 45) = 14.12$, $p < 0.001$, with a relatively strong effect (partial $\eta^2 = 0.239$). This points to the continued importance of prior ability in shaping results and highlights why it needs to be considered in the analysis. Although some variability in scores remains unexplained ($SS = 4310.23$; $df = 45$; $MSE = 95.78$), a substantial portion is captured by the model.

However, students in the experimental group achieved higher levels of critical thinking than those in the control group, even after accounting for their initial differences. Beyond the statistical findings, these results carry broader implications. Strengthening critical thinking skills can improve how individuals approach problems and make decisions, which in turn supports productivity and reduces the need for further training. In the longer term, such gains contribute to the development of human capital and may foster innovation and efficiency within the wider economy.

The adjusted post-test scores for the control and experimental groups are summarized in Table 7. After accounting for initial pre-test differences, students in the control group achieved an adjusted mean score of 53.2, with a standard error of 1.7 and a 95% confidence interval between 49.8 and 56.6. Meanwhile, the experimental group attained a higher adjusted mean of 60.8, with a standard error of 1.8 and a 95% confidence interval ranging from 57.1 to 64.5. These adjusted means, which reflect post-intervention performance while controlling for baseline ability, suggest that students who engaged with the integrated assessment approach performed better in critical thinking compared to those who received traditional assessments.

Table 7: Estimated marginal means

Group	Adjusted mean	Standard error	95% confidence interval	Interpretation
Control group	53.2	1.7	[49.8, 56.6]	Lower critical thinking skills
Experimental group	60.8	1.8	[57.1, 64.5]	Higher critical thinking skills (due to intervention)

Note: The 95% confidence intervals do not overlap substantially, providing further evidence of a true difference between groups

Importantly, the confidence intervals for the two groups exhibit very limited overlap, with the upper bound of the control group (56.6) falling below the lower bound of the experimental group (57.1). Although close, this gap implies a distinct difference in performance that supports the statistical findings from the ANCOVA. The lack of substantial overlap strengthens the evidence that the observed differences are not due to random variation but rather attributable to the impact of the intervention. In conclusion, the estimated marginal means analysis provides additional support for the effectiveness of the integrated assessments in enhancing students' critical thinking capabilities beyond what could be expected from initial skill levels alone.

Figure 3 illustrates the adjusted mean post-test critical thinking scores for both the control and experimental groups, taking into account differences in initial (pre-test) performance. The experimental group, which participated in the integrated assessment intervention, attained a noticeably higher adjusted mean score (60.8) compared to the control group (53.2). The vertical lines represent the standard errors associated with each mean. The gap between the two bars visually reinforces the statistical significance observed in the ANCOVA, highlighting the positive impact of the integrated assessment approach on students' critical thinking development.

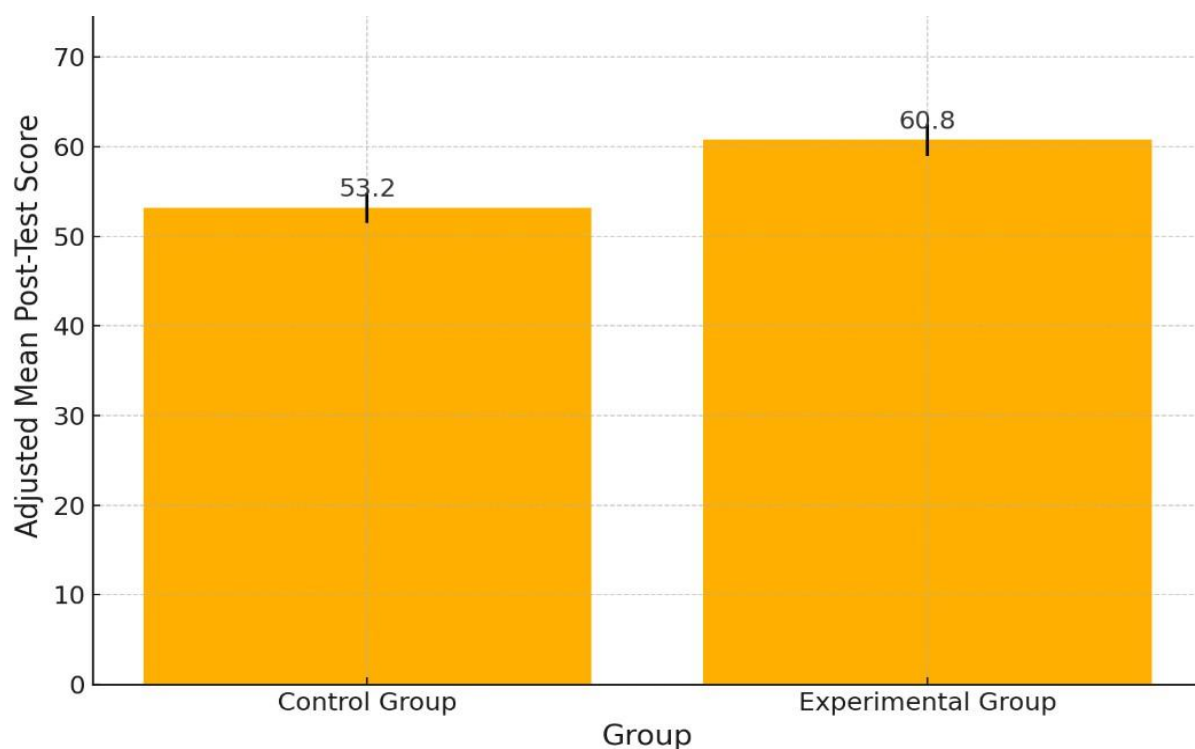


Figure 3: Adjusted mean post-test scores by group

Understanding whether the intervention leads to meaningful improvements beyond baseline differences is essential for evaluating its effectiveness. Table 8 presents a comparison of the adjusted post-test scores for the experimental and control groups after accounting for initial performance.

Table 8: Adjusted group comparison and effect size

Comparison	Mean difference	SE	95% CI lower	95% CI upper	Cohen's <i>d</i>
Experimental vs control	7.6	2.48	2.61	12.59	0.78

The experimental group scored, on average, 7.6 points higher than the control group. The confidence interval (95% CI [2.61, 12.59]) does not include zero, indicating that this difference is unlikely to have occurred by chance. The effect size (Cohen's $d = 0.78$) suggests a moderate to strong impact, highlighting that the improvement is not only statistically reliable but also practically meaningful. In practical terms, this indicates that the intervention led to a noticeable enhancement in students' critical thinking skills. These findings provide strong support for the effectiveness of the integrated assessment approach in improving learning outcomes, even after accounting for students' initial abilities.

5. Discussion

The findings of this study provide strong empirical support for the effectiveness of integrated assessments in enhancing critical thinking skills among first-year accounting students. The significantly higher post-test scores observed in the experimental group align with prior research emphasizing the value of integrated and authentic assessment approaches. These results are consistent with the findings of Vanderheide and Walkington (2008), who reported that integrated, multidisciplinary assessments promote deeper learning and improve cognitive engagement. Recent studies by Quinlan et al. (2025) and Villarroel et al. (2018) emphasize that integrated assessment fosters meaningful learning and enhances students' ability to apply knowledge across contexts.

Similarly, McPhun (2010) found that integrated assessments enhance students' preparedness for the workplace, although their study relied primarily on perception-based data. This finding is reinforced by contemporary research, such as Jackson (2019) and Sokhanvar et al. (2021), which shows that integrated and authentic assessment enhances employability skills, particularly in terms of workplace readiness, problem-solving, and the ability to transfer knowledge into practice. In contrast, the present study extends this body of knowledge by providing quantitative evidence of improvements in critical thinking using the Watson-Glaser instrument. Furthermore, the results support the theoretical propositions of Paul and Nosich (2013), who argue that effective assessments should require knowledge transfer, interdisciplinary thinking, and real-world problem-solving. This position is reinforced by more recent studies, such as Chamba and Chikusvura (2024), which demonstrate that integrated assessment systems enable deeper learning and the development of competencies required for complex, real-world contexts. The integrated assessment used in this study incorporated these elements through scaffolded, real-world tasks, which likely contributed to the observed gains in critical thinking performance.

The findings also reinforce the arguments by De Klerk et al. (2025), who emphasize that integrated assessments enhance higher-order thinking by breaking down

disciplinary silos and promoting interconnected learning. The improved performance of the experimental group suggests that such integration allows students to engage more meaningfully with content, thereby fostering analytical and evaluative thinking skills. In contrast, the lower performance of the control group supports earlier critiques of traditional assessment methods (Huba & Freed, 2000; Lombard & Grosser, 2008), which tend to encourage surface learning rather than deep cognitive engagement. This highlights the limitations of conventional assessments in developing critical thinking competencies. Overall, this study not only confirms existing literature but also contributes new empirical evidence within the South African higher education context, demonstrating that integrated assessments are an effective pedagogical strategy for developing critical thinking skills in accounting education.

6. Conclusion and Recommendations

The primary objective of this study was to examine the relationship between integrated assessments and the development of critical thinking skills among first-year accounting students at a UoT in South Africa. The findings provide empirical evidence that contextualized integrated assessments, particularly those that simulate real-world scenarios, significantly enhance students' critical thinking performance. Students exposed to such assessments achieved higher scores on standardized measures, including the Watson-Glaser instrument, compared to those assessed using traditional methods. This study makes an important contribution to the literature by addressing a notable gap in accounting education research, where the development of critical thinking is often discussed conceptually but less frequently measured using quantitative, standardized instruments. By demonstrating measurable improvements in critical thinking, the findings extend prior theoretical work, including that of Paul and Nosich (2013), which emphasizes the importance of knowledge transfer, interdisciplinary thinking, and real-world problem-solving in effective assessment design.

From a practical perspective, the study highlights that transforming assessment practices in higher education requires a shift not only in instructional design but also in institutional culture. While integrated assessments offer clear pedagogical benefits, their development is often complex and time-intensive. Academics, as subject-matter experts, may not always have formal training in teaching and assessment design. Therefore, institutional support structures, particularly centers for teaching and learning, play a critical role in equipping educators with the necessary skills to design and implement integrated assessments that effectively foster critical thinking. Furthermore, the findings emphasize the importance of a collaborative, multi-stakeholder approach. Effective integrated assessment design requires alignment between program outcomes, teaching practices, and assessment strategies. Collaboration among academics across modules enhances curriculum coherence, while engagement with industry partners and professional bodies ensures that assessments reflect authentic workplace demands. These partnerships are essential in developing graduates who possess not only technical competence but also the critical thinking skills required in complex, real-world environments.

Notwithstanding its contributions, this study has several limitations. The research was conducted in a single institution and focused on first-year accounting students, which may limit the generalizability of the findings. Additionally, while the Watson–Glaser instrument provides a robust measure of critical thinking, future studies could incorporate mixed-method approaches to capture deeper insights into student learning experiences. Future research should explore the longitudinal impact of integrated assessments on critical thinking development across different levels of study and disciplines. Comparative studies across institutions and contexts would further strengthen the evidence base. In addition, further investigation into the design principles and scalability of integrated assessments would be valuable in supporting broader implementation.

In conclusion, this study provides strong empirical support for the use of contextualized integrated assessments as a means of developing critical thinking in accounting education. It is anticipated that these findings will inform and enhance assessment practices within UoTs and the wider higher education sector, contributing to the development of graduates who are better equipped to navigate the complexities of the modern workplace.

6.1 Limitations

The data collection instrument and sample size were limitations of the current study. Although the W-GCTA has a long history and has been used in numerous countries and settings, including South Africa, an instrument designed specifically for the South African context would have been more appropriate. From the literature reviewed, it was evident that such a culturally appropriate critical thinking test is not available at present in South Africa (Lombard & Grosser, 2008). It is hoped that further research will be conducted in this area and that a critical thinking test will be designed specifically for the South African context. While the sample size was modest, the results provide valuable insights for curriculum designers. Future research with larger, randomized samples is recommended to confirm these findings.

6.2 Suggested Areas for Further Research

It is recommended that integrated assessments based on real-world problems should be conducted in the first and subsequent years. It is not always possible for technology universities to offer cooperative education/in-service training to all students. Therefore, contextualized integrated assessments bridge this gap by allowing students to experience workplace requirements without physically being in the work environment.

7. References

- Amirova, N. (2025). Traditional vs. non-traditional teaching in secondary education: A comparative analysis. *Porta Universorum*, 1(3), 101–109. <https://doi.org/10.69760/portuni.010309>
- Barac, K., & Du Plessis, L. (2014). Teaching pervasive skills to South African accounting students. *Southern African Business Review*, 18(1), 53–79. <https://doi.org/10.25159/1998-8125/5645>
- Bernard, R. M., Zhang, D., Abrami, P. C., Sicol, F., Borokhovski, E., & Surkes, M. A. (2008). Exploring the structure of the Watson–Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3(1), 15–22.
- Bhardwaj, V., Zhang, S., Tan, Y. Q., & Pandey, V. (2025). Redefining learning: Student-centered

- strategies for academic and personal growth. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1518602>
- Bernard, R. M., Zhang, D., Abrami, P. C., Sicol, F., Borokhovski, E., & Surkes, M. A. (2008). Exploring the structure of the Watson-Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity*, 3(1), 15-22.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation Centre for Excellence in Learning and Teaching*. (2014). *Assessment policy*. Durban University of Technology.
- Chamba, L. T., & Chikusvura, N. (2024). Future-proofing quality education using integrated assessment systems. *Quality Education for All*, 1(1), 240-255. <https://doi.org/10.1108/QEA-11-2023-0014>
- Cloete, M. (2018). The impact of an integrated assessment on the critical thinking skills of first-year university students. *Accounting Education*, 27(5), 479-494. <https://doi.org/10.1080/09639284.2018.1501717>
- Connolly, D., Dickinson, L., & Hellewell, L. (2023). The development of undergraduate employability skills through authentic assessment in college-based higher education. *Journal of Learning Development in Higher Education*, 27. <https://doi.org/10.47408/jldhe.vi27.1004>
- De Klerk, C., Ker-Fox, J., & Steenekamp, L. (2025). Enhancing critical thinking through collaborative learning: The impact of a partial pre-release assessment format. *Accounting Education*, 34(4), 499-532. <https://doi.org/10.1080/09639284.2024.2361647>
- Dewey, J. (1997). *How We Think*. Dover Publications, Inc.: Mineola, New York.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational leadership*, 43(2), 44-48.
- Facione, P. A. (2011). Critical thinking: What it is and why it counts. *Insight assessment*, 1(1), 1-23.
- Hansson, S. O. (2019). *Critical thinking*. Wiley.
- Hassan, K. E., & Madhum, G. (2007). Validating the Watson Glaser critical thinking appraisal. *Higher Education*, 54(3), 361-383.
- Huba, M. E., & Freed, J. E. (2000). *Comparison of teacher-centered and learner-centered paradigms*. <http://assessment.uconn.edu/docs/TeacherCenteredVsLearnerCenteredParadigms.pdf>
- Jackson, D. (2019). Student perceptions of the development of employability skills through part-time work education. *Education + Training*, 61(4), 473-489. <https://doi.org/10.1108/ET-10-2018-0216>
- Jackson, D., Michelson, G., & Munir, R. (2023). Developing accountants for the future: New technology, skills, and the role of stakeholders. *Accounting Education*, 32(2), 150-177.
- Kroon, N., do Céu Alves, M., & Martins, I. (2021). The impacts of emerging technologies on accountants' role and skills: Connecting to open innovation – A systematic literature review. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(3), Article 163. <https://doi.org/10.3390/joitmc7030163>
- Ku, K. Y. L. (2009). Assessing student critical thinking performance: Urging for measurement using multi-response format. *Thinking Skills and Creativity*, 4(1), 70-76. <https://doi.org/10.1016/j.tsc.2009.02.001>
- Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's research reports*, 6(1), 40-41.
- Lombard, B. J. J. (2008). Modelling critical thinking through learner-oriented assessment. *South African Journal of Higher Education*, 22(5), 1029-1043. <https://journals.co.za/doi/pdf/10.10520/EJC37485>
- Lombard, K., & Grosser, M. (2008). Critical thinking: are the ideals of OBE failing us or are we failing the ideals of OBE? *South African journal of education*, 28(4), 561-579.
- McPhun, H. (2010). *Integrated Assessments: Engaging Ways to Enhance Learner Outcomes*. AKO. <https://ako.ac.nz/assets/Knowledge-centre/Integrated-assessments-Engaging-ways-to-enhance-learner-outcomes/Integrated-assessments-Engaging-ways-to-enhance-learner-outcomes.pdf>
- Mehdizade, N. (2025). The advantages and disadvantages of teacher-centred instruction in foreign language teaching amidst contemporary demands: An analysis based on the experience of Azerbaijan. *EuroGlobal Journal of Linguistics and Language Education*, 2(5), 119-131. <https://doi.org/10.69760/egille.2505010>
- Nwosu, L. I., & Vorster, H. (2021). A systematic literature review on critical thinking in an undergraduate accounting programme. *International Journal of Financial Research*, 12(4), 225.

- Organisation for Economic Co-operation and Development. (2019). *OECD Learning Compass 2030: A series of concept notes*.
https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/1-1-learning-compass/OECD_Learning_Compass_2030_Concept_Note_Series.pdf
- Paul, R., & Nosich, G. M. (2013). A model for the national assessment of higher order thinking. *Foundation for Critical Thinking*. <https://www.criticalthinking.org/pages/a-model-for-the-national-assessment-of-higher-order-thinking/591>
- Quinlan, K. M., Sellei, G., & Fiorucci, W. (2025). Educationally authentic assessment: Reframing authentic assessment in relation to students' meaningful engagement. *Teaching in Higher Education*, 30(3), 717-734.
- Reichardt, C. S. (2009). Quasi-experimental design. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 46-71). SAGE Publications
- Saifuddin, T., Yousuf, U., Jabeen, S., & Matloob, F. (2025). A comparative study of student-centered interactive teaching strategies and traditional approaches in history education. *Policy Research Journal*, 3(3), 372-383. <https://doi.org/10.5281/zenodo.15043632>
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (8th ed.). Pearson.
- Shaw, A., Liu, O. L., Gu, L., Kardonova, E., Chirikov, I., Li, G., Hu, S., Yu, N., Ma, L., & Guo, F. (2020). Thinking critically about critical thinking: validating the Russian HEIghten® critical thinking assessment. *Studies in Higher Education*, 45(9), 1933-1948.
- Sin, S., Jones, A., & Wang, Z. (2015). Critical thinking in professional accounting practice: Conceptions of employers and practitioners. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 431-456). Springer.
- Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in educational evaluation*, 70, 101030.
- South African Institute of Professional Accountants. (2019, September 16). *In the 4IR era, professional accountants must develop critical thinking*. <https://www.saipa.co.za/in-the-4ir-era-professional-accountants-must-develop-critical-thinking/>
- South African Qualifications Authority. (2006). *Investigating the use of critical cross-field outcomes in the design of ABET qualifications and unit standards: Terms of reference*. South African Qualifications Authority.
- Terblanche, E., & De Clercq, B. (2021). A critical thinking competency framework for accounting students. *Accounting Education*, 30(4), 325-354
- Vanderheide, R. A., & Walkington, J. (2008). Maximised learning through integrated assessment: Evidenced through nursing practice. *Nurse Education in Practice*, 9(6), 351-353. <https://doi.org/10.1016/j.nepr.2008.10.010>
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840-854.
- Watson, G., & Glaser, E. (2002). *Watson-Glaser Critical Thinking Appraisal: UK Edition manual*. The Psychological Corporation.
- Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review*, 109(4), 21-32. <https://doi.org/10.3200/AEPR.109.4.21-32>