

Playing the devil's advocate: A pathway to stronger critical thinking skills

Samah Ahmad ¹ , Samia Shahpo ¹ , Alaaeldin Hamid ² , Abdelrahim Fathy Ismail ^{2*} 

¹ College of Sciences and Humanities - Jubail, Imam Abdulrahman Bin Faisal University, Dammam, SAUDI ARABIA

² Department of Curriculum and Instruction, College of Education, King Faisal University, Al Ahsa 31982, SAUDI ARABIA

*Corresponding Author: afismail@kfu.edu.sa

Citation: Ahmad, S., Shahpo, S., Hamid, A., & Ismail, A. F. (2026). Playing the devil's advocate: A pathway to stronger critical thinking skills. *Pedagogical Research*, 11(4), em0296. <https://doi.org/10.29333/pr/19632>

ARTICLE INFO

Received: 10 Jun. 2026

Accepted: 10 Sep. 2026

ABSTRACT

Saudi Arabian college students may face challenges in developing critical thinking skills, which are considered essential for the future workforce under Saudi Arabia's Vision 2030 initiative. The present study used a quasi-experimental design to test the effectiveness of the Devil's Advocate approach in improving critical thinking among Saudi undergraduates. Forty-three students enrolled in two equivalent course sections were assigned to an experimental group (n = 22) employing the Devil's Advocate method or a control group (n = 21) engaging in traditional consensus-oriented discussion across a 15-week semester. Critical thinking was assessed at pretest and posttest using the California Critical Thinking Skills Test Form 2000 (CCTST; Insight Assessment, 2013), yielding scores on five reasoning domains (Analysis, Inference, Evaluation, Induction, and Deduction) and an Overall composite score. Independent-samples t-tests confirmed baseline equivalence (all $p > .65$). Within the experimental group, paired-samples t-tests revealed statistically significant and very large gains on every domain (Cohen's $d = 1.64-2.15$, all $p < .001$), whereas the control group showed only small, non-significant changes. ANCOVA controlling for pretest scores confirmed that the experimental group significantly outperformed the control group on every outcome, with large effect sizes (partial $\eta^2 = .324-.445$). The intervention shifted the median experimental participant from the Moderate to the Strong CCTST proficiency tier, $\chi^2(3, N = 43) = 15.69$, $p = .001$, Cramér's $V = .604$. Non-parametric tests corroborated every parametric inference. These findings support cognitive conflict, social constructivist, and metacognitive accounts of how structured dissent advances reasoning, and demonstrate that the Devil's Advocate approach is a low-cost, theoretically grounded intervention for addressing critical thinking deficits in Saudi higher education and supporting the human-capital aspirations of Vision 2030. Practically, these findings indicate that a low-cost, easily embedded classroom protocol can yield substantial critical-thinking gains, offering instructors, curriculum designers, and higher-education policymakers a scalable, resource-light strategy for strengthening students' reasoning without wholesale curricular reform.

Keywords: devil's advocate, critical thinking, CCTST, quasi-experimental, Saudi higher education, higher-order thinking skills

INTRODUCTION

Critical thinking is a foundational competency for academic learning, professional practice, and twenty-first-century citizenship. Critical thinking is a multifaceted cognitive process that involves analyzing information, drawing justified inferences, evaluating evidence, and applying both inductive and deductive reasoning to reach well-supported conclusions (Facione, 2015; Halpern, 2014; Moon, 2008). Together, these abilities form the foundation for reasoned analysis and decision-making in academic and real-world contexts.

The cultivation of critical thinking has become one of the central aims of higher education worldwide. Universities are increasingly expected to graduate students who can evaluate evidence, weigh competing arguments, and make reasoned judgements in conditions of uncertainty rather than simply reproduce transmitted knowledge. Employers, accreditation bodies, and international frameworks for twenty-first-century skills consistently rank critical thinking among the most valued graduate attributes, yet a substantial body of research indicates that conventional lecture-based and consensus-oriented instruction does comparatively little to advance these higher-order competencies. This gap between the stated priority of critical thinking and its uneven cultivation in practice has motivated growing interest in active, dialogic, and argument-based pedagogies that deliberately engage students in reasoning.

Among these pedagogies, the Devil's Advocate approach is notable for its explicit focus on structured dissent as a driver of reasoning. By requiring learners to construct and defend positions contrary to a prevailing view, the technique compels the

analysis of counterarguments, the evaluation of assumptions, and the generation of alternative inferences—precisely the cognitive operations that define critical thinking. Because it can be embedded within existing course content without additional materials or curricular restructuring, the approach is well suited to higher-education settings seeking practical, low-cost means of strengthening students' reasoning. Establishing the educational significance of this approach provides the foundation for examining its value in the Saudi context, to which the discussion now turns.

In the Kingdom of Saudi Arabia (KSA), the development of strong critical thinking skills has increasingly become a major national concern. Indeed, as part of the Saudi Vision 2030 initiative, the need for the development of a transformational educational framework that produces graduates who can analyze, evaluate and innovate in fields including engineering, medicine, and business is evident (Alharbi, 2022; Alhowail & Albaqami, 2024; Al-Shakhis & Banks-Santilli, 2023). However, in spite of this strategic requirement, studies have found the presence of a notable weakness in the ability of Saudi students to engage effectively in critical thinking. This weakness is primarily attributed to the culture of memorization and teacher-centric instruction in schools and universities that does not encourage students to think critically, argue rationally and challenge assumptions (Al-Zahrani & Elyas, 2017; Allamnakhrah, 2013; Khawaji, 2024). Hence, many students at Saudi universities find themselves unable to demonstrate their competencies in critical thinking tests that examine skills of inference and evaluation of assumptions and claims (Alhowail & Albaqami, 2024), which negatively affects employability where critical thinking is consistently rated by employers as an essential competency (Alrashedi, 2025).

The challenge is compounded by a curriculum that has not fully integrated critical thinking across disciplines and a documented gap in teacher preparation for effectively delivering and assessing critical thinking content (Al-Zahrani & Elyas, 2017; Alharbi, 2022). Despite some recent reforms that have included critical thinking in the curriculum, the use of critical thinking in the classroom is still underdeveloped, as shallow understanding is still considered better than deep reflection (Aldossari & Aldajani, 2021). Thus, there is a dire need for interventions that will allow Saudi students to actively engage themselves in analyzing and evaluating information and counter-arguing it. This study focuses on fulfilling the aforementioned requirement by investigating the impact of the Devil's Advocate approach on critical thinking abilities of Saudi university students.

LITERATURE REVIEW

Definition and Historical Background

The Devil's Advocate approach is a method in which an individual deliberately assumes a contrary position or expresses an opposing viewpoint in a discussion, regardless of personal beliefs, with the aim of challenging prevailing opinions and stimulating critical thinking (Shewring, 2024). The technique encourages rigorous examination of assumptions, evidence, and arguments by introducing dissent and alternative perspectives, helping to uncover potential flaws and biases (Mann et al., 2024). The idea of Devil's Advocate arose during the sixteenth century in the Roman Catholic Church, where an *advocatus diaboli* was assigned to oppose the canonization of saints for a thorough examination process (Blumberg, 2017). The modern use cases of the Devil's Advocate role can be found not only in law but also in such spheres as education, intelligence assessment, and organizational decision-making; in those cases, the method is used as a means of controlling and ensuring more flexible conclusions (Leal et al., 2023). Besides, the method exploits various psychological effects, such as the confirmation bias effect (Deeb et al., 2018).

Theoretical Framework

Three theoretical approaches underpin the application of the Devil's Advocate approach in critical-thinking education. According to the cognitive conflict theory, a learner's knowledge acquisition is provoked by discrepancies in their perception, resulting in cognitive disequilibrium (Kang et al., 2004; Piaget, 1985). In this respect, the Devil's Advocate prompts cognitive disequilibrium and supports deep analysis rather than rote memorization of facts (Davis, 2013). Further, according to the social constructivist theory, knowledge is socially constructed within collaborative environments, where learning is guided and scaffolded by more competent learners (Powell & Kalina, 2009; Prawat, 1992). Consistent with social constructivism, the Devil's Advocate is a means of socially constructing knowledge via dialogical collaboration and the introduction of productive dissent into discussions (Schweizer et al., 2002); the technique represents a practical example of Vygotskian zones of proximal development, as the assigned advocate surpasses the analytical abilities of the entire group and pulls other participants along (Lehtinen et al., 2023; Vygotsky, 1978). Finally, metacognition theory stresses learners' monitoring of their cognitive activity, as well as self-regulation thereof (Flavell, 1979). Students who are assigned an opposite opinion by a teacher and tasked with its defense and critical analysis monitor and regulate their reasoning process deliberately (Davis, 2013; Deeb et al., 2018), which fosters self-awareness and strategic flexibility essential for lifelong learning (New South Wales Department of Education, 2024).

Benefits and Pedagogical Use

Three related benefits linked with Devil's Advocate have been identified in relation to education. The first benefit is the development of skills that enable learners to analyze counterarguments. Through analyzing and understanding counterarguments, learners develop a better capacity to critically analyze other positions, think about evidence, identify biases and develop flexible thinking that will prepare them to make decisions and debate in a democratic society (Davis, 2013; Gay, 2021). Another benefit of this practice is improved argumentation and problem-solving skills as the practice helps learners build an argument, consider the counterargument to that argument, and deal with issues using flexibility (Bicksler & Hannah, 2022; Davis, 2013; Manchang et al., 2022). Lastly, deeper cognitive processing is enhanced through the need to think about counterarguments which prevents the problem of groupthink and fosters intellectual flexibility (Deeb et al., 2018; Shewring, 2024).

Successful classroom implementation entails strategically allocating learners the responsibility of arguing the opposing view, switching roles, implementing it through debates and engaging in Socratic questions that necessitate the provision of evidence on every viewpoint (Walker, 2004). This strategy can be included in discussion protocol or peer tutoring where students can take initiative and engage actively (DeTemple et al., 2026). Reflection after the session further improves metacognition skills (Fogarty, 2009). From research, it is clear that the Devil's Advocate role increases engagement, participation and helps develop students' cognitive abilities regardless of whether it is a physical class or an online environment (Benjamin & Schwartz, 1994; Bicksler & Hannah, 2022; Dobbs, 2025).

The Devil's Advocate and Critical Thinking

Critical thinking refers to the set of cognitive skills and dispositions required to reason effectively, solve problems, and make decisions. Some of the most commonly identified skills include analysis, interpretation, evaluation, inference, explanation, and self-regulation or metacognition (Ennis, 2011; Facione, 1990; Halpern, 2014). The technique called Playing Devil's Advocate involves a strategic cognitive exercise aimed at improving critical thinking by forcing oneself to analyze perspectives that are opposite to one's viewpoint. This strategic practice explicitly tackles confirmation bias because the individual is made to look for flaws and alternative perspectives rather than seeking supportive arguments. Institutionalization of dissent transforms System 1 cognition based on heuristics into the more demanding System 2 cognition (Halpern, 2014) and thus helps develop intellectual humility and the need for a variety of criteria (Paul & Elder, 2008). In organizational settings, structured dissent can lead to better quality decisions through exposure to flawed assumptions, an understanding of the problem's totality, and reduced overconfidence (Schwenk, 1989).

The literature shows empirical support for the use of the Devil's Advocate strategy in minimizing the occurrence of groupthink, increasing psychological safety, and improving the quality of decision-related reasoning (Klocke, 2007; Lunenburg, 2012; Nemeth et al., 2001; Postmes et al., 2001; Schwenk, 1989). Similarly, classrooms studies show that students taking on the role of Devil's Advocate critically evaluate their assumptions and arguments which improves their logical reasoning skills (Akhmad et al., 2021; Davis, 2013; Manchang et al., 2022; Zenker, 2014). Moreover, recent scholarly literature provides an application of the strategy in modern contexts such as the area of moral philosophy (Chouhan, 2025) and decision making in workplace safety situations (Galloway, 2025). Although the evidence base for the application of the technique is generally positive, few empirical studies have been undertaken in Saudi Arabia where lack of critical thinking skills among students is a known issue. Existing literature from Saudi Arabian contexts mentions system-related barriers such as learning via rote memorization and teacher-centrism; however, little research has been conducted on novel approaches that focus on using cognitive conflict to enhance metacognition (Al Ghamdi & Deraney, 2013; Algouzi et al., 2023). There are very few studies on the use of Devil's Advocate approach in Saudi Arabian classroom settings especially ones utilizing standardized measures of critical thinking skills. The current study attempts to fill in the research gap by investigating the effects of the strategy on critical thinking skills of Saudi Arabian university students.

In sum, the reviewed literature establishes three points. First, structured dissent—and the Devil's Advocate technique in particular—is theoretically grounded in cognitive-conflict, social-constructivist, and metacognitive accounts of learning. Second, empirical work across organizational and classroom settings consistently links the technique to reduced groupthink and improved reasoning. Third, and most importantly, this evidence base is drawn almost entirely from Western and East Asian contexts and rarely employs standardized, psychometrically validated measures of critical thinking. Within Saudi higher education—where rote learning and teacher-centred instruction are documented barriers to critical thinking—no experimental study has yet tested the Devil's Advocate approach using a standardized instrument. The present study addresses this gap by evaluating the effect of a structured Devil's Advocate intervention on the critical thinking of Saudi university students, as measured by the CCTST. Accordingly, three research questions guided this study:

- RQ1:** To what extent does a 15-week Devil's Advocate intervention improve the overall critical thinking performance of Saudi college students, as measured by the California Critical Thinking Skills Test Form 2000 (CCTST)?
- RQ2:** Does the intervention produce differential effects across the five CCTST Form 2000 reasoning domains (Analysis, Inference, Evaluation, Induction, and Deduction)?
- RQ3:** How does the experimental group's gain compare with that of a control group engaged in traditional consensus-oriented discussion of the same content?

METHODOLOGY

Research Design and Participants

The study adopted a quasi-experimental pre-test/post-test design involving a control group where two equivalent classes were assigned randomly to the experimental and control groups. The study involved 43 undergraduate students from Imam Abdulrahman bin Faisal University, Saudi Arabia, who were attending two equivalent classes of one course instructed by the same instructor within the period of one 15-week academic semester. Of these, one class made up the Experimental Group ($n = 22$) and the other made up the Control Group ($n = 21$). The sample size was determined by practical rather than a priori power considerations: the two intact course sections available during the study semester constituted the entire accessible population for the course, and whole classes were retained to preserve ecological validity and to avoid contamination between the experimental and control conditions. A post-hoc sensitivity analysis indicated that the achieved sample ($N = 43$) provided adequate power ($\geq .80$, $\alpha = .05$, two-tailed) to detect large effects ($d \geq 0.80$), while acknowledging limited power to detect small-to-moderate effects. This constraint is revisited in the Limitations section.

Table 1. Score ranges and proficiency categories of the CCTST Form 2000

Domain	Score Range	Proficiency Categories
Analysis	0 - 7	Not Manifested (0-2), Moderate (3-4), Strong (≥ 5)
Inference	0 - 16	Not Manifested (0-5), Moderate (6-11), Strong (≥ 12)
Evaluation	0 - 11	Not Manifested (0-3), Moderate (4-7), Strong (≥ 8)
Induction	0 - 17	Not Manifested (0-5), Moderate (6-11), Strong (≥ 12)
Deduction	0 - 17	Not Manifested (0-5), Moderate (6-11), Strong (≥ 12)
Overall	0 - 34	Not Manifested (0-7), Weak (8-12), Moderate (13-18), Strong (19-23), Superior (≥ 24)

Note. Score ranges and proficiency categories are taken directly from the CCTST Test Manual (Insight Assessment, 2013), **Tables 3** and **4**. The subscale system uses three proficiency levels; the Overall score uses five levels. Items contribute to multiple subscales simultaneously; subscale scores therefore do not sum to the Overall score.

Instrument

The California Critical Thinking Skills Test Form 2000 (CCTST; Insight Assessment, 2013) was used to measure critical thinking. The CCTST is a psychometrically validated, paper-and-pencil, multiple-choice instrument designed for undergraduate and graduate students. Form 2000 consists of 34 items and assesses five core reasoning domains—Analysis, Inference, Evaluation, Induction, and Deduction—yielding both subscale scores and an Overall composite score. The test is discipline-neutral and has demonstrated adequate internal consistency ($KR-20 = .78-.82$) and predictive validity for academic and professional outcomes (Wheeler & Collins, 2003). **Table 1** presents the score ranges and proficiency categories of the CCTST Form 2000. The CCTST was administered in its original English version. English-language proficiency was not separately assessed or statistically controlled in the present study; as most participants were non-native English speakers, the potential influence of language comprehension on test performance is acknowledged as a limitation (see Section *Limitations*).

Procedure

The intervention was conducted over a full 15-week academic semester. In the experimental section, one participant was formally assigned the Devil's Advocate role for each case analysis. The assigned student was required to deliberately challenge prevailing viewpoints by presenting counterarguments and alternative perspectives, regardless of personal beliefs. The Devil's Advocate role was systematically rotated so that each student served in the role at least once during the semester, ensuring that the intervention was a class-wide experience rather than one concentrated in a subset of participants. To reinforce metacognitive reflection, all experimental-group students completed a structured journal entry after each session documenting how the Devil's Advocate arguments had influenced their reasoning and assumptions. The control group engaged with identical course content through traditional consensus-oriented discussions, which emphasized agreement and justification of preferred solutions without structured dissent. Both groups received the same instructional materials and were taught by the same instructor to control for teaching effects. The CCTST was administered to both groups in the first and final weeks of the semester. Because the intervention was embedded in an essay-writing course, the Devil's Advocate sessions centred on controversial topics relevant to the Saudi context about which students first developed essays and a prevailing class position. Representative examples included whether social media does more harm than good for Saudi youth; whether English should replace Arabic as the medium of instruction in Saudi universities; and whether the rapid expansion of entertainment and tourism under Vision 2030 strengthens or erodes Saudi cultural identity. For each topic, the assigned Devil's Advocate was required to construct and defend the position opposing the emerging class consensus. Each session followed a standardized sequence in which the class first articulated a prevailing position, after which the assigned Devil's Advocate presented structured counterarguments before the discussion was opened to the group. To ensure that students in the Devil's Advocate role engaged in genuine structured dissent rather than superficial opposition, the instructor monitored role adherence during each session against a brief checklist (requiring at least two evidence-based counterarguments and at least one explicit challenge to an underlying assumption), and the post-session reflective journals were reviewed to confirm substantive engagement with the opposing position. To standardize delivery across conditions and reduce teacher-expectancy effects, the same instructor followed identical lesson plans, time allocations, materials, and assessment criteria in both sections and refrained from signalling to students which instructional condition was expected to be more effective.

Data Analysis

Data were analyzed using IBM SPSS Statistics with $\alpha = .05$ (two-tailed). Descriptive statistics summarised pretest and posttest performance. Independent-samples t-tests established baseline equivalence, paired-samples t-tests assessed within-group change, and analysis of covariance (ANCOVA) compared posttest scores while controlling for the corresponding pretest score. Effect sizes (Cohen's d for paired comparisons; partial eta squared, η^2p , for ANCOVA) quantified the magnitude of observed differences. All parametric tests were corroborated with non-parametric alternatives (Wilcoxon signed-rank for within-group; Mann-Whitney U for between-group). A chi-square test of independence with Cramér's V compared the distribution of students across CCTST proficiency tiers between groups.

RESULTS

The analysis proceeded in four stages: (a) preliminary assumption checks and baseline equivalence; (b) descriptive analysis of pretest and posttest performance; (c) within-group paired-samples t-tests; and (d) between-group ANCOVA on posttest scores controlling for pretest performance, complemented by effect-size estimates and a categorical analysis of CCTST proficiency tiers.

Table 2. Independent-samples t-tests of pretest CCTST scores

Domain	EG M (SD)	CG M (SD)	Mean Diff.	t (41)	p	Cohen's d
Overall	13.45 (2.84)	13.81 (3.06)	-0.35	-0.39	.695	-0.12
Analysis	2.95 (1.13)	3.10 (1.18)	-0.14	-0.40	.692	-0.12
Inference	6.18 (1.74)	6.43 (1.83)	-0.25	-0.45	.653	-0.14
Evaluation	3.41 (1.30)	3.52 (1.44)	-0.11	-0.28	.785	-0.08
Induction	6.86 (1.91)	7.05 (2.04)	-0.18	-0.31	.761	-0.09
Deduction	5.05 (1.68)	5.19 (1.75)	-0.15	-0.28	.783	-0.08

Note. df = 41 for all comparisons. p values are two-tailed. All $|d| \leq 0.14$, indicating negligible baseline differences.

Table 3. Descriptive statistics of CCTST performance at pretest and posttest by group

Domain	Group	Pretest M (SD)	Posttest M (SD)	Mean Gain	95% CI of Gain	Cohen's d
Overall	EG (n = 22)	13.45 (2.84)	19.18 (3.13)	5.73	[4.18, 7.28]	1.64
	CG (n = 21)	13.81 (3.06)	14.95 (3.17)	1.14	[-0.11, 2.40]	0.41
Analysis	EG (n = 22)	2.95 (1.13)	4.55 (1.10)	1.59	[1.17, 2.02]	1.66
	CG (n = 21)	3.10 (1.18)	3.38 (1.20)	0.29	[-0.22, 0.79]	0.26
Inference	EG (n = 22)	6.18 (1.74)	9.32 (1.89)	3.14	[2.49, 3.78]	2.15
	CG (n = 21)	6.43 (1.83)	7.05 (1.94)	0.62	[-0.06, 1.30]	0.41
Evaluation	EG (n = 22)	3.41 (1.30)	6.18 (1.50)	2.77	[2.15, 3.40]	1.96
	CG (n = 21)	3.52 (1.44)	4.00 (1.52)	0.48	[-0.18, 1.13]	0.33
Induction	EG (n = 22)	6.86 (1.91)	9.77 (2.07)	2.91	[2.14, 3.68]	1.67
	CG (n = 21)	7.05 (2.04)	7.62 (2.16)	0.57	[-0.13, 1.27]	0.37
Deduction	EG (n = 22)	5.05 (1.68)	7.68 (1.81)	2.64	[1.97, 3.30]	1.76
	CG (n = 21)	5.19 (1.75)	5.71 (1.85)	0.52	[-0.16, 1.21]	0.35

Note. EG = Experimental Group (Devil's Advocate); CG = Control Group (traditional discussion). Cohen's d for paired-samples comparisons, computed as mean gain divided by SD of gain. CCTST Form 2000 score ranges per Insight Assessment (2013): Analysis 0-7, Inference 0-16, Evaluation 0-11, Induction 0-17, Deduction 0-17, Overall 0-34.

Preliminary Analyses and Baseline Equivalence

Shapiro-Wilk tests indicated that the Overall CCTST score and the Evaluation, Induction, and Deduction subscales were normally distributed in both groups at both time points (all $p > .05$). The Inference subscale showed marginal departures from normality in two of four cells ($p = .042$ and $.047$), and the Analysis subscale showed more pronounced non-normality across all four cells (W values 0.761-0.901; p values $< .001$ to $.032$), consistent with the subscale's narrow score range (0-7) and the resulting floor effect at pretest. To address this, all parametric findings for Analysis and Inference were corroborated with non-parametric alternatives, which yielded the same pattern of inference at every test. Given the documented robustness of t and ANCOVA to non-normality at moderate sample sizes with balanced groups (Blanca et al., 2017) and the convergent non-parametric results, parametric analyses were retained as primary.

Levene's test confirmed homogeneity of variance across groups for every domain at both time points (pretest $F = 0.05$ - 0.47 , all $p > .49$; posttest $F < 0.01$ - 0.76 , all $p > .38$). The ANCOVA assumption of homogeneity of regression slopes was met (Group $\times$ Pretest $F = 0.00$ - 0.87 , all $p \geq .356$). Independent-samples t-tests on pretest scores established baseline equivalence: no statistically significant differences emerged between groups on any CCTST domain (all $|t| < 0.46$, all $p > .65$; all $|d| \leq 0.14$, negligible by Cohen's (1988) standards; see [Table 2](#)).

Descriptive Statistics

Descriptive statistics, including means, standard deviations, mean gain scores, 95% confidence intervals for gains, and within-group Cohen's d, are summarised in [Table 3](#). At pretest, the EG mean Overall CCTST score ($M = 13.45$, $SD = 2.84$) and the CG mean ($M = 13.81$, $SD = 3.06$) both fell within the lower Moderate proficiency category (13-18; Insight Assessment, 2013), indicating limited baseline critical thinking competence in both groups. This pattern is consistent with previous reports of weak baseline critical thinking performance among Saudi tertiary students (Alharbi, 2022; Al Ghamdi & Deraney, 2013).

At posttest, the EG advanced substantially to a mean Overall CCTST score of 19.18 ($SD = 3.13$), placing it within the Strong category, while the CG showed only a modest increase to 14.95 ($SD = 3.17$), remaining within the Moderate category. The same divergent pattern was observed across all five subscales: EG mean gains ranged from 1.59 (Analysis) to 3.14 (Inference), whereas CG gains ranged from 0.29 to 0.62. The 95% CIs of the EG gains excluded zero in every domain, whereas the corresponding CG intervals included zero throughout, indicating that the control group's gains were not statistically reliable.

Within-Group Pretest-to-Posttest Comparisons

Paired-samples t-tests within each group ([Table 4](#)) confirmed that the experimental group exhibited statistically significant gains on every CCTST domain (all $p < .001$). The largest within-group t value occurred on Inference, $t(21) = 10.10$, $p < .001$, $d = 2.15$, followed by Evaluation, $t(21) = 9.21$, $p < .001$, $d = 1.96$; Deduction, $t(21) = 8.26$, $p < .001$, $d = 1.76$; Induction, $t(21) = 7.83$, $p < .001$, $d = 1.67$; Analysis, $t(21) = 7.78$, $p < .001$, $d = 1.66$; and the Overall composite, $t(21) = 7.69$, $p < .001$, $d = 1.64$. All EG effect sizes were very large by Cohen's (1988) standards ($d > 0.80$). Wilcoxon signed-rank tests confirmed the EG gains across every domain (Overall: $z = -4.06$, $p < .001$; Analysis: $z = -3.82$, $p < .001$; Inference: $z = -4.11$, $p < .001$; Evaluation: $z = -4.01$, $p < .001$; Induction: $z = -3.92$, $p < .001$; Deduction: $z = -4.01$, $p < .001$), and confirmed the non-significant pattern for the CG (all $p > .08$).

Table 4. Paired-samples t-tests of CCTST pretest-to-posttest gains within each group

Domain	Group	Mean Gain	SD of Gain	t	p	Cohen's d
Overall	EG	5.73	3.49	7.69	< .001	1.64
	CG	1.14	2.76	1.90	.072	0.41
Analysis	EG	1.59	0.96	7.78	< .001	1.66
	CG	0.29	1.10	1.19	.249	0.26
Inference	EG	3.14	1.46	10.10	< .001	2.15
	CG	0.62	1.50	1.89	.073	0.41
Evaluation	EG	2.77	1.41	9.21	< .001	1.96
	CG	0.48	1.44	1.52	.144	0.33
Induction	EG	2.91	1.74	7.83	< .001	1.67
	CG	0.57	1.54	1.71	.104	0.37
Deduction	EG	2.64	1.50	8.26	< .001	1.76
	CG	0.52	1.50	1.60	.126	0.35

Note. df = 21 for EG and 20 for CG. p values are two-tailed. Cohen's d computed as mean gain divided by SD of gain (within-subjects).

Table 5. ANCOVA results comparing posttest CCTST scores between groups after adjusting for pretest performance

Domain	Adj. M (EG)	Adj. M (CG)	F (1, 40)	p	Partial η^2	95% CI of MD
Overall	19.27	14.86	26.10	< .001	.395	[2.66, 6.15]
Analysis	4.59	3.34	19.19	< .001	.324	[0.67, 1.82]
Inference	9.41	6.96	32.07	< .001	.445	[1.58, 3.33]
Evaluation	6.21	3.97	31.52	< .001	.441	[1.44, 3.06]
Induction	9.84	7.55	22.68	< .001	.362	[1.32, 3.26]
Deduction	7.73	5.66	22.77	< .001	.363	[1.19, 2.94]

Note. Adj. M = estimated marginal mean controlling for pretest score; MD = adjusted mean difference (EG – CG). Effect-size magnitude based on Cohen's (1988) benchmarks for partial η^2 (small $\geq .01$, medium $\geq .06$, large $\geq .14$).

Table 6. Distribution of participants across CCTST overall proficiency categories

Time / Group	Weak n (%)	Moderate n (%)	Strong n (%)	Superior n (%)	Median Tier	Total n
Pretest – EG	8 (36.4)	13 (59.1)	1 (4.5)	0 (0.0)	Moderate	22
Pretest – CG	8 (38.1)	12 (57.1)	1 (4.8)	0 (0.0)	Moderate	21
Posttest – EG	1 (4.5)	5 (22.7)	13 (59.1)	3 (13.6)	Strong	22
Posttest – CG	6 (28.6)	12 (57.1)	3 (14.3)	0 (0.0)	Moderate	21

Note. Proficiency categories based on CCTST Form 2000 Overall-score cut-offs (Insight Assessment, 2013): Not Manifested (0-7), Weak (8-12), Moderate (13-18), Strong (19-23), Superior (≥ 24). Percentages are computed within each row.

Within the control group, in contrast, no comparison reached statistical significance. The largest CG comparison was marginal (Overall: $t(20) = 1.90$, $p = .072$, $d = .41$); all subscale comparisons yielded p values between .073 and .249 and small-to-moderate effect sizes ($d = .26$ -.41). Wilcoxon tests confirmed the non-significant pattern. The juxtaposition of uniformly significant, very-large within-group effects in the EG against small, non-significant changes in the CG provides initial evidence for the differential impact of the Devil's Advocate intervention.

Between-Group Comparisons: ANCOVA

One-way ANCOVAs were conducted for each CCTST outcome, with group as the fixed factor, posttest score as the dependent variable, and the corresponding pretest score as the covariate (Table 5). After controlling for pretest performance, the experimental group significantly outperformed the control group on every CCTST domain (all $p < .001$), with large effect sizes (partial $\eta^2 = .324$ -.445). On the Overall CCTST, EG adjusted M = 19.27 versus CG adjusted M = 14.86, $F(1, 40) = 26.10$, $p < .001$, partial $\eta^2 = .395$, 95% CI of MD [2.66, 6.15]; approximately 39.5% of the variance in posttest performance not explained by the covariate was attributable to the intervention.

The strongest between-group effects emerged for Inference, $F(1, 40) = 32.07$, $p < .001$, partial $\eta^2 = .445$ (95% CI of MD [1.58, 3.33]), and Evaluation, $F(1, 40) = 31.52$, $p < .001$, partial $\eta^2 = .441$ (95% CI of MD [1.44, 3.06]). Robust large effects were also found for Deduction (partial $\eta^2 = .363$), Induction (partial $\eta^2 = .362$), and Analysis (partial $\eta^2 = .324$). All partial η^2 values substantially exceeded the .14 threshold for a large effect (Cohen, 1988). Mann-Whitney U tests on posttest scores corroborated all ANCOVA conclusions (Overall: $U = 382$, $p < .001$; Analysis: $U = 348$, $p = .003$; Inference: $U = 368$, $p < .001$; Evaluation: $U = 392$, $p < .001$; Induction: $U = 352$, $p = .003$; Deduction: $U = 360$, $p = .002$), with rank-biserial correlations ranging from .50 to .70.

Distribution of Proficiency Categories

Students were classified into CCTST Overall proficiency tiers at both time points using the cut-off scores specified in the CCTST Test Manual (Insight Assessment, 2013); the distribution is shown in Table 6. At pretest, both groups performed within the lower Moderate tier (EG: 8 Weak, 13 Moderate, 1 Strong; CG: 8 Weak, 12 Moderate, 1 Strong), with EG and CG pretest means of 13.45 and 13.81 respectively. At posttest, only 4.5% of the EG remained Weak, while 72.7% had advanced to Strong (59.1%) or Superior (13.6%). In contrast, 28.6% of the CG remained Weak at posttest and none reached Superior. A chi-square test of independence confirmed that the posttest distribution differed significantly between groups, $\chi^2(3, N = 43) = 15.69$, $p = .001$, Cramér's $V = .604$, representing a large association.

Summary of Findings

Five findings converge across all analytical layers. First, groups were statistically equivalent at baseline, both predominantly in the lower Moderate tier of the CCTST. Second, the experimental group showed statistically significant and very large within-group gains on every CCTST domain (all $p < .001$; $d = 1.64-2.15$), while the control group showed only small, non-significant changes ($d = .26-.41$). Third, ANCOVA confirmed that the experimental group significantly outperformed the control group on every domain after pretest adjustment (large partial $\eta^2 = .324-.445$). Fourth, the intervention shifted the majority of EG participants into the Strong/Superior proficiency tiers (Cramér's $V = .604$). Fifth, non-parametric tests corroborated every parametric conclusion, attesting to robustness. The strongest effects were observed for Inference and Evaluation, with comparably robust but slightly smaller effects on the remaining subscales.

DISCUSSION

This study aimed to test whether the implementation of a 15-week Devil's Advocate training program would increase the critical thinking skills of the Saudi college students. The findings showed statistically significant and educationally meaningful improvements in the scores on Overall CCTST and all its five subscales for the experimental group, whereas the control group showed minor improvement which was not significant. These findings were still robust after conducting non-parametric tests, and there was a shift upwards among the experimental participants from Moderate Proficiency to Strong and Superior. The following sections of this discussion will analyze the findings based on three theoretical frameworks, contextualize the results within prior literature, and consider their implications

Theoretical Interpretation

The findings were consistent with all three theoretical perspectives used in this study. As far as cognitive conflict theory is concerned (Kang et al., 2004; Piaget, 1985), the Devil's Advocate approach successfully instantiated cognitive disequilibrium in a regular and consistent way: weekly, each student was required to present and argue for the opposite position to the one accepted as true in the group, which required the class to consider contrary information. Particularly notable was the very large within-groups effect size ($d = 1.96$) and ($d = 2.15$) observed on Evaluation and Inference, the two subscales most directly linked to making judgments about conflicting information and drawing appropriate conclusions—a clear instantiation of cognitive conflict theory predictions. On the other hand, the consensus-based control condition intentionally reduced cognitive conflict, producing small effect sizes ($d = .26-.41$), which confirms the idea that it was the cognitive conflict embedded in the Devil's Advocate protocol, rather than discussion per se, that drove the observed improvement.

In terms of social constructivism (Powell & Kalina, 2009; Vygotsky, 1978), the Devil's Advocate can be considered as a scaffold which draws particular attention to the social construction of meaning through the formal designation of one person as responsible for representing alternative viewpoints. As a result, the procedure provides the participants with a publicly accessible artefact of dissent which allows for the co-construction of knowledge as structured peer discourse as noted by Lehtinen et al. (2023). The fact that the gains extended far beyond the confines of the argument-oriented subscales and involved both inductive and deductive reasoning indicates that the scaffolding applied to the process of reasoning generalized to both reasoning skills with the designated Devil's Advocate temporarily exceeding the class's overall capabilities and bringing the rest of the group members closer to their zone of proximal development. From the perspective of metacognition (Davis, 2013; Deeb et al., 2018; Flavell, 1979), the structured reflective journals can be viewed as metacognitive monitoring and self-regulation instruments. In line with the metacognitive framework, the experimental group performed better in both deductively oriented (Deduction; partial $\eta^2 = .363$) and more broadly evaluative tasks (Evaluation; partial $\eta^2 = .441$). The transition to the Superior tier as measured by Cramér's $V (.604)$ indicates a qualitative shift in how students approached thinking, rather than a quantitative increment within an unchanged repertoire.

Relation to Prior Empirical Research

The results of the study are consistent with previous research on the effects of the Devil's Advocate technique. Specifically, the meta-analysis carried out by Schwenk (1989) and more recently the studies conducted by Nemeth et al. (2001) and Postmes et al. (2001) have revealed the efficacy of structured dissent in improving the quality and quantity of arguments during decision-making processes. In extending the research to the Saudi tertiary educational setting where there has been a relative dearth of studies on the subject, this study has provided effect size statistics based on data collected using a standardized, internationally-calibrated scale. Results of this study are in harmony with the findings of a Devil's Advocate study conducted in secondary schools by Manchang et al. (2022), as well as the qualitative findings of a Devil's Advocate study by Davis (2013).

With respect to the Saudi context, the baseline CCTST performance observed here (both groups performing within the lower Moderate tier at pretest) quantitatively confirms qualitative concerns voiced in prior research (Alharbi, 2022; Alhowail & Albaqami, 2024; Al-Zahrani & Elyas, 2017). The pretest means (EG = 13.45, CG = 13.81), both within the lower Moderate category (13-18), are consistent with prior reports that Saudi tertiary students enter higher education with under-developed analytic dispositions (Allamnakhrah, 2013; Khawaji, 2024).

Differential Effects across Subscales and Pedagogical Implications

While effect sizes were equally large across all five subscales of the CCTST, the largest between-group effects were found in Inference (partial $\eta^2 = .445$) and Evaluation (partial $\eta^2 = .441$). This pattern is expected theoretically because the Devil's Advocate technique entails the immediate and repeated use of these exact processes in order to construct an argument against a given

point of view, as the assigned advocate has to produce new inferences out of contentious assumptions and evaluate the legitimacy of the dominant viewpoint. Deduction (partial $\eta^2 = .363$) and Induction (partial $\eta^2 = .362$) showed comparably large but slightly smaller effects, consistent with their role as tools the advocate uses in service of inference and evaluation. The smaller effect size in Analysis (.324) can be at least partially accounted for by the narrow range of the subscale (0-7) and the presence of the floor effect in pretest; the within-group Cohen's d of 1.66 indicates that the practical impact was, in fact, comparable to other subscales when scale constraints are considered.

The magnitude of the observed effects (within-group $d = 1.64$ – 2.15 ; partial $\eta^2 = .324$ – $.445$) is large relative to typical educational-intervention benchmarks and therefore warrants cautious interpretation. Several features of the present study plausibly contributed to these gains. Participants began from a low baseline within the lower Moderate tier, leaving considerable room for measurable growth and reducing ceiling constraints. The intervention combined several potentially reinforcing components—formal role assignment, systematic role rotation, and mandatory post-session reflective journaling—delivered intensively over a full 15-week semester, which may have produced compounding effects greater than any single element in isolation. Delivery by a single, committed instructor in two intact sections, together with the novelty of the approach in a context where structured dissent is uncommon, may also have amplified the observed differences. These considerations suggest that the effect sizes reported here should be regarded as an upper-bound estimate obtained under favourable conditions; comparable magnitudes should not be assumed in larger, multi-site, or multi-instructor implementations, and replication is needed before the findings are generalized.

These findings suggest three implications for higher education. Firstly, the large size of the effect suggests that there is no need for systemic change to achieve significant improvements in critical thinking skills. The Devil's Advocate protocol represents a highly focused and inexpensive intervention that does not involve any additional material resources and any changes to existing courses' curricula, thus being in line with Aldossari and Aldajani's (2021) suggestion about how the difference between the policy and its implementation in Saudi higher education can be bridged through embedded course-level interventions. Secondly, the consistent improvement seen on all five subscales suggests that the Devil's Advocate approach fosters a more generalizable disposition rather than the development of skills that are relevant only to certain situations. Indeed, although students were trained to take the position of dissenter in regard to course content, the improvements observed were relevant to discipline-neutral critical thinking measures, thus supporting the assertion that training in dialogical skills is indeed transferable (Holman, 2000; Wu, 2026). Finally, the Devil's Advocate method may prove particularly suitable for contexts where open intellectual dissent is uncommon. Specifically, the protocol assigns the role of dissenter to a particular student on a specific day of the week, thus decoupling disagreement from risk associated with personal confrontation, a phenomenon that has been reported by Al-Zahrani and Elyas (2017) and Khawaji (2024).

Although the study was conducted in Saudi Arabia, its implications extend to higher-education settings more broadly. The mechanisms through which the Devil's Advocate approach operates—externalizing dissent, depersonalizing disagreement, and prompting metacognitive reflection—are not culturally specific, and the intervention is discipline-neutral and readily embedded within existing courses. The approach may therefore be particularly useful in any educational context characterized by consensus-oriented norms, reluctance to voice disagreement, or limited resources for wholesale curricular reform, while institutions with more established cultures of debate may adopt it as a structured complement to existing active-learning practices. At the same time, because the present findings derive from a single institution and cultural setting, cross-context generalization should be treated as a hypothesis for future comparative and multi-site research rather than an established conclusion.

Limitations

Two limitations qualify the current findings. First, the sample pool was derived from one institution and one course alone; therefore, there is a lack of external generalizability. Second, a relatively small number of participants ($N=43$) sufficed for detecting the extremely large effect sizes noted but lacked sufficient statistical power for detecting smaller effect sizes; more participants could enable such detection and also make possible moderator analysis (such as baseline proficiency and gender differences). Beyond these, several additional limitations should be noted. Because both sections were taught by the same instructor, individual teaching style was held constant, but this design cannot rule out instructor-related threats to internal validity, including differential enthusiasm, subconscious teaching toward the assessed critical-thinking domains in the experimental section, and novelty effects; although standardized lesson plans and role-adherence monitoring were used to mitigate teacher-expectancy effects, no formal blinding of the instructor was possible. Relatedly, assignment occurred at the classroom-section level ($k = 2$) while the statistical analyses treated individual students ($N = 43$) as independent units; sample-size constraints made multilevel modelling of this nested structure infeasible, and the resulting clustering may lead to an underestimation of standard errors. The intervention was also relatively brief (a single 15-week semester), and no delayed or follow-up assessment was administered, so the durability and long-term transfer of the observed gains remain unknown. Finally, the CCTST was administered in English to predominantly non-native English speakers without a separate measure or control of language proficiency, which may have introduced construct-irrelevant variance. These considerations, together with the small single-institution sample and the unusually large effect sizes, reinforce the need for cautious interpretation and for larger, multi-site, and longitudinal replication.

Implications

The findings carry theoretical, educational, and practical implications. Theoretically, the convergent gains across all five reasoning domains provide empirical support, within a previously under-studied context, for the integrated cognitive-conflict, social-constructivist, and metacognitive account of how structured dissent advances reasoning; the results demonstrate that these mechanisms can be operationalized in an authentic classroom and produce measurable change on a standardized instrument.

Educationally, the study indicates that critical thinking among Saudi university students is malleable rather than fixed and can be strengthened through a structured, course-embedded protocol that requires no additional materials or curricular restructuring. This positions the Devil's Advocate approach as a feasible instructional tool for instructors seeking to foster higher-order thinking within content-oriented courses, and it underscores the value of pairing role rotation with reflective journaling to depersonalize disagreement and promote metacognition.

Practically and at the level of policy, the results suggest that institutional and national goals for graduate reasoning skills—including the human-capital aspirations articulated in Vision 2030—can be advanced through low-cost, scalable classroom interventions complemented by regular standardized assessment of critical thinking. Faculty development programmes could incorporate training in structured-dissent techniques, and institutions could use standardized measures such as the CCTST to monitor progress. These implications should nonetheless be weighed against the study's limitations and confirmed through larger and more diverse replications.

CONCLUSION AND RECOMMENDATIONS

The present study explored the effect of the Devil's Advocate on the critical thinking skills of students enrolled in universities in Saudi Arabia. This is an effort to address the lack of critical thinking skills among Saudi tertiary education students, which is in line with one of the goals of the vision of Saudi Arabia 2030, where the aim is to develop graduates that think independently and innovatively. Using a quasi-experimental pretest/posttest control group design, with the CCTST as a standardized instrument, the study compared a Devil's Advocate condition with a consensus discussion condition in two equivalent course sections at Imam Abdulrahman bin Faisal University.

Findings demonstrated that the Devil's Advocate produced educationally relevant as well as statistically significant outcomes regarding the Overall CCTST composite scale as well as each of its five subscales. Within the experimental group, there was a considerable increase in test scores (Cohen's $d = 1.64$ - 2.15 , all $p < .001$) compared to almost no improvement for the control group. Using ANCOVA, it was found that the experimental group significantly outperformed the control group on all measures when controlling for the pretest differences (partial $\eta^2 = .324$ -. $.445$). According to the categorical analysis, the treatment caused the change in the median experimental participant's score from the lower tier of the Moderate classification to the Strong tier. All findings derived through non-parametric statistical methods confirmed parametric ones as well. Taken together, these results present empirical proof (in a sphere, where this kind of evidence was lacking before) that the critical thinking problem among Saudi university students is not their inherent, unchangeable feature but is susceptible to educational interventions. In particular, by combining the elements of cognitive conflict, social-constructivist scaffolding, and metacognitive reflection in one relatively cost-free procedure, the Devil's Advocate provides a theoretically founded and practically implementable method for developing the critical thinking abilities that Vision 2030 considers as essential to the Kingdom's future workforce.

Recommendations for Practice

Based on the findings of the current study, the following practice-based implications are provided:

- Instructors should incorporate an effective Devil's Advocate protocol within content-oriented classes in which one of the course goals is related to fostering critical thinking skills. The Devil's Advocate protocol would involve students rotating the role of devil's advocate followed by reflective journal writing after each discussion.
- Vision 2030 can be achieved by integrating the use of Devil's Advocate as a course-based initiative that can complement curricular reform since it can be introduced into any course without changing its content as an attempt to achieve the objectives of Vision 2030.
- Instructors training regarding depersonalizing strategies such as formalized role allocation, rotation, and reflective journals can play a central role in helping them implement the Devil's Advocate in situations where there is no open debate culture in classrooms.
- Institutional assessment frameworks should supplement the assessment of course-based critical thinking by conducting regular standardized tests of critical thinking skills using tools like CCTST, allowing comparable tracking of progress against Vision 2030 benchmarks.

Recommendations for Future Research

Four possible directions for future research are suggested:

- Multi-site replication using larger samples in order to calculate the magnitude of effects accurately and establish the generalizability of findings to different universities and fields of expertise.
- Use of delayed post-test methodology to measure the durability of effects over long periods of time and their usefulness in actual problem solving situations beyond the CCTST.
- Mixed methods research combining CCTST results with the analysis of reflections and classroom discussions in order to understand the mechanisms that underlie the effectiveness of the Devil's Advocate technique.
- Comparative research comparing the Devil's Advocate with other structured dialogue techniques such as dialectical inquiry, Socratic seminar, and formal debates.

Author contributions: **SZA:** conceptualization, supervision; **AF:** methodology, formal analysis, interpretation, writing – original draft; **SMS & AAH:** data curation, implementation, interpretation, writing – review & editing. All authors reviewed and approved the final manuscript.

Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Proposal Number: KFU263140).

Acknowledgements: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Proposal Number: KFU263140).

Ethical statement: The authors stated that participants were informed about the purpose of the study and their role in it. Participation was voluntary, and informed consent was obtained from all participants before data collection. Ethical approval was granted by the Research Ethics Committee at King Faisal University (Approval No. KFU-2026-ETHICS4922- 09-01-2026). The collected data were stored securely and used only for the purposes of this study.

AI statement: The authors stated that ChatGPT and Grammarly were used only to improve the language and readability of the manuscript. The authors reviewed all outputs and take full responsibility for the content of the manuscript.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy and ethical considerations related to the participants.

REFERENCES

- Akhmad, M., Chang, S., & Deguchi, H. (2021). Closed-mindedness and insulation in groupthink: Their effects and the devil's advocacy as a preventive measure. *Journal of Computational Social Science*, 4(2), 455–478. <https://doi.org/10.1007/s42001-020-00083-8>
- Aldossari, A. T., & Aldajani, M. M. (2021). The effectiveness of a self-questioning strategy at developing academic achievement and critical-thinking skills among secondary-school students in Saudi Arabia. *International Journal of Learning, Teaching and Educational Research*, 20(8), 278–299. <https://doi.org/10.26803/ijlter.20.8.17>
- Algouzi, S., Alzubi, A. A. F., & Nazim, M. (2023). Enhancing EFL students' critical thinking skills using a technology-mediated self-study approach: EFL graduates and labor market in perspective. *PLOS ONE*, 18(10), Article e0293273. <https://doi.org/10.1371/journal.pone.0293273>
- Al Ghamdi, A. K. H., & Deraney, P. M. (2013). Effects of teaching critical thinking to Saudi female university students using a stand-alone course. *International Education Studies*, 6(7), 176–188. <https://doi.org/10.5539/ies.v6n7p176>
- Alharbi, M. M. (2022). *Investigating the potential of blended learning to develop critical thinking skills among student teachers in Saudi Arabia* [Doctoral dissertation, University of Leicester]. University of Leicester Research Repository.
- Alhowail, A. M., & Albaqami, S. E. (2024). Evaluation of the critical thinking skills of secondary school students in Saudi Arabia. *International Journal of Research in Education and Science*, 10(1), 16–35.
- Allamnakhrah, A. (2013). Learning critical thinking in Saudi Arabia: Student perceptions of secondary pre-service teacher education programs. *Journal of Education and Learning*, 2(1), 197–210. <https://doi.org/10.5539/jel.v2n1p197>
- Alrashedi, A. K. (2025). Future workforce skills for Saudi Arabia: A PLS-SEM analysis of Vision 2030 readiness. *Information Development*. Advance online publication. <https://doi.org/10.1177/02666669251357311>
- Al-Shakhs, W. M., & Banks-Santilli, L. (2023). Promoting critical thinking skills in students in middle and high school to achieve the Kingdom of Saudi Arabia's National Vision 2030. *Working Papers in Social Sciences Research Journal*, 6(1), 77–88.
- Al-Zahrani, R. S., & Elyas, T. (2017). The implementation of critical thinking in a Saudi EFL context: Challenges and opportunities. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 1(2), 41–55. <https://doi.org/10.21093/ijeltal.v1i2.21>
- Benjamin, S., & Schwartz, W. (1994). When less is more: A Devil's Advocate position on standards. *English Journal*, 83(5), 28–30. <https://doi.org/10.58680/ej19947519>
- Bicksler, W. H., & Hannah, P. (2022). The Devil's Advocate role in asynchronous online discussions: Asian region undergraduate perspectives. *Issues in Educational Research*, 32(1), 36–56.
- Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29(4), 552–557. <https://doi.org/10.7334/psicothema2016.383>
- Blumberg, J. (2017, December). The psychological science of persuasion: Presenting your case to overcome jurors' built-in resistance. *Advocate Magazine*.
- Chouhan, R. (2025). The Devil's Advocate in moral philosophy: Contradiction as a path to truth. *International Journal of English and Studies*, 7(7), 397–399. <https://doi.org/10.47311/IJOES.2025.18.07.399>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203771587>
- Davis, J. R. (2013). Improving students' critical thinking and classroom engagement by playing the Devil's Advocate. *Theory, Research and Action in Urban Education*, 11(1).
- Deeb, H., Vrij, A., Hope, L., Mann, S., Leal, S., Granhag, P. A., & Strömwall, L. A. (2018). The Devil's Advocate approach: An interview technique for assessing consistency among deceptive and truth-telling pairs of suspects. *Legal and Criminological Psychology*, 23(1), 37–52. <https://doi.org/10.1111/lcrp.12114>
- DeTemple, J., Hayes, H. E., & West, J. E. (2026). *The dialogic classroom in higher education: Revolutionary listening for curiosity, engagement, and deep learning*. Taylor & Francis.

- Dobbs, M. (2025). *Bespoke ELA: Innovative lessons to spark student engagement*. John Wiley & Sons.
- Ennis, R. H. (2011). *The nature of critical thinking: An outline of critical thinking dispositions and abilities*. University of Illinois.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. American Philosophical Association.
- Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Fogarty, R. J. (Ed.). (2009). *Brain-compatible classrooms* (3rd ed.). Corwin Press.
- Galloway, S. M. (2025). Seeking dissenting views to shape successful safety decisions. *Professional Safety*, 70(2), 14–17.
- Gay, S. E. K. P. (2021). A method to develop critical thinking: Playing the Devil's Advocate. In D. Koyama (Ed.), *Development of innovative pedagogical practices for a modern learning experience* (pp. 269–292). CSMFL Publications. <https://doi.org/10.46679/978819484836310>
- Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). Psychology Press. <https://doi.org/10.4324/9781315885278>
- Holman, D. (2000). A dialogical approach to skill and skilled activity. *Human Relations*, 53(7), 957–980. <https://doi.org/10.1177/0018726700537003>
- Insight Assessment. (2013). *CCTST test manual*. California Academic Press.
- Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education*, 34, 71–96. <https://doi.org/10.1023/B:RISE.0000020885.41514.41>
- Khawaji, A. (2024). Attainment of critical thinking aims through the English curriculum: A study from Jeddah, Saudi Arabia. *Pakistan Journal of Life and Social Sciences*, 22(1), 475–494. <https://doi.org/10.57239/PJLSS-2024-22.1.0034>
- Klocke, U. (2007). How to improve decision making in small groups: Effects of dissent and training interventions. *Small Group Research*, 38(3), 437–468. <https://doi.org/10.1177/1046496407301974>
- Leal, S., Vrij, A., Deeb, H., Dabrowna, O., & Fisher, R. P. (2023). Combining the Devil's Advocate approach and the verifiability approach to assess veracity in opinion statements. *European Journal of Psychology Applied to Legal Context*, 15(2), 53–61. <https://doi.org/10.5093/ejpalc2023a6>
- Lehtinen, A., Kostiainen, E., & Näykki, P. (2023). Co-construction of knowledge and socioemotional interaction in pre-service teachers' video-based online collaborative learning. *Teaching and Teacher Education*, 131, Article 104299. <https://doi.org/10.1016/j.tate.2023.104299>
- Lunenburg, F. (2012). Devil's advocacy and dialectical inquiry: Antidotes to groupthink. *International Journal of Scholarly Academic Intellectual Diversity*, 14(2), 1–9.
- Manchang, B., Padung, L. D., & Pofung, D. D. (2022). Developing secondary school students' critical thinking through Devil's Advocate approach in the classroom. *Journal of African Social Studies*, 3(2), 35–44.
- Mann, S., Vrij, A., Deeb, H., & Leal, S. (2024). All mouth and trousers? Use of the Devil's Advocate questioning protocol to determine authenticity of opinions about protester actions. *Psychiatry, Psychology and Law*, 31(5), 909–931. <https://doi.org/10.1080/13218719.2023.2208063>
- Moon, J. A. (2008). *Critical thinking: An exploration of theory and practice*. Routledge.
- Nemeth, C. J., Brown, K. S., & Rogers, J. D. (2001). Devil's Advocate versus authentic dissent: Stimulating quantity and quality of argument. *European Journal of Social Psychology*, 31(6), 707–720. <https://doi.org/10.1002/ejsp.58>
- New South Wales Department of Education. (2024). *Metacognition: A key to unlocking learning*. New South Wales Government.
- Paul, R., & Elder, L. (2008). *The miniature guide to critical thinking concepts and tools*. Foundation for Critical Thinking Press.
- Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development* (T. Brown & K. J. Thampy, Trans.). University of Chicago Press. (Original work published 1975)
- Postmes, T., Spears, R., & Cihangir, S. (2001). Quality of decision making and group norms. *Journal of Personality and Social Psychology*, 80(6), 918–930. <https://doi.org/10.1037/0022-3514.80.6.918>
- Powell, K. C., & Kalina, C. J. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, 130(2), 241–250.
- Prawat, R. S. (1992). Teachers' beliefs about teaching and learning: A constructivist perspective. *American Journal of Education*, 100(3), 354–395. <https://doi.org/10.1086/444021>
- Schweizer, H., Whipp, J., & Hayslett, C. (2002). Quality control in online courses: Using a social constructivist framework. *Computers in the Schools*, 19(3–4), 143–158. https://doi.org/10.1300/J025v19n03_12
- Schwenk, C. R. (1989). A meta-analysis of the comparative effectiveness of devil's advocacy and dialectical inquiry. *Strategic Management Journal*, 10(3), 303–306. <https://doi.org/10.1002/smj.4250100309>
- Shewring, E. (2024). The application of the devil's advocacy technique to intelligence analysis. *National Security Journal*, 6(2). <https://doi.org/10.36878/nsj20240925.03>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

- Walker, S. A. (2004). Socratic strategies and devil's advocacy in synchronous CMC debate. *Education + Training*, 46(6/7), 301–309. <https://doi.org/10.1108/00400910410555204>
- Wheeler, L. A., & Collins, S. K. R. (2003). The influence of concept mapping on critical thinking in baccalaureate nursing students. *Journal of Professional Nursing*, 19(6), 339–346. [https://doi.org/10.1016/S8755-7223\(03\)00134-7](https://doi.org/10.1016/S8755-7223(03)00134-7)
- Wu, Y. (2026). Enhancing student critical thinking skills and dispositions through a multi-agent system with argumentation and Devil's Advocacy. *Educational Technology & Society*, 29(2), 339–354. [https://doi.org/10.30191/ETS.202604_29\(2\).SP07](https://doi.org/10.30191/ETS.202604_29(2).SP07)
- Zenker, F. (2014). Know thy biases! Bringing argumentative virtues to the classroom. In D. Mohammed & M. Lewiński (Eds.), *Virtues of argumentation: Proceedings of the 10th OSSA Conference* (pp. 1–13). University of Windsor.