OPEN ACCESS
Received: 28 -10 -2025
Accepted: 05- 01-2025

الآداب
للدراسات اللغوية والأدبية

Critical Thinking and Holistic English Language Development at the University of Ha'il: Advancing Vision 2030 Competencies

Dr. Shirien Gaffar Mohammed Abdulrahim *

Shirien@uoh.edu.sa

Abstract

This research looks into how well critical thinking teaching works when paired with full-spectrum English learning at the University of Hail in Saudi Arabia. For phase one, around 100 young men joined a trial that checked growth in logical reasoning - measured through the California Critical Thinking TEST CCTST - and their grasp of academic English using Academic English Proficiency Test AEPT scores. Instead of random groups, researchers ran a pre-and-post comparison backed by Analysis of Covariance ANCOVA math; after this number crunching came follow-up talks with a dozen learners and teachers to help make sense of the data patterns. The numbers showed a clear, big boost in CT scores (d=1..49), along with solid progress in Holistic English Language Development HELD skills (d=0.89) within the test group. Better thinking abilities came together with sharper language use - backed up by a strong link between them ($r = 0.68$, $p < 0.001$). Students' comments hinted at more interest and confidence while learning this way. Yet, their insights also highlighted problems like unprepared instructors or rigid systems that slowed down long-term adoption. The findings show clear proof that CT speeds up learning a second language, fitting well with Vygotsky's idea of social learning. Because of this link, colleges at the University of Hail ought to prioritize blended teaching strategies. Instead of sticking to old ways, combining methods offers a tested way to build flexible thinkers. By doing so, UoH moves closer to meeting goals outlined in Saudi Arabia's Vision 2030 plan.

Keywords: Thinking skills, Critical thinking, Learning environment, Hybrid education, Career readiness.

* Assistant Professor of Applied Linguistics, Department of English Language, Arts and Humanities, University of Hail, Saudi Arabia.

Cite this article as: Abdulrahim, S. J. M. (2026). Critical Thinking and Holistic English Language Development at the University of Ha'il: Advancing Vision 2030 Competencies, *Arts for Linguistic & Literary Studies*, *8*(1): 586 -602. https://doi.org/10.53286/15rx7402

الآداب
للدراسات اللغوية والأدبية

OPEN ACCESS
تاريخ الاستلام: 2025/10/28م
تاريخ القبول: 2025/01/05م

التفكير النقدي والتطوير الشامل لمهارات اللغة الإنجليزية في جامعة حائل: تعزيز كفايات رؤية 2030

د. شيرين جعفر محمد عبدالرحيم *
Shirien@uoh.edu.sa

الملخص

تبحث هذه الدراسة في فاعلية تعليم مهارات التفكير النقدي عند ربطها بتنمية كفاءة اللغة الإنجليزية في جامعة حائل بالسعودية. بدأت العملية بتجربة شملت نحو 100 طالب؛ حيث جرى تقييم التقدم في الاستدلال المنطقي عبر استخدام اختبارCCTST، في حين تم قياس الكفاية اللغوية الأكاديمية باختبار .AEPT لم يعتمد البحث على عينة عشوائية، بل استخدم تصميماً قبليا-بعديا مدعوماً بتحليل ANCOVA إحصائياً. بعد ذلك، أجريت مقابلات مع 12 مشاركًا من الطلاب والمعلمين لكشف دوافع النتائج وفهم سياقاتها بدقة. كانت هناك زيادة واضحة في مهارات التفكير النقدي عند المجموعة التي تتبع الطريقة الجديدة (d = 1.49) بينما سُجل تحسنٌ متزامن في المهارات المتعلقة باللغة (HELD) بدرجة أقل لكنها مهمة (d = 0.89). علاوة على ذلك، أشارت النتائج إلى أن من يحققون تقدماً أكبر في التفكير المنطقي غالباً ما يكون أداؤهم الأفضل لغوياً (r = 0.68, p < 0.001). من جهة أخرى، أفاد المتعلمون بأن حماسهم زاد وبأنهم شعروا بطلاقة أكثر خلال هذه التجربة. رغم هذا، كشفت آراؤهم عن صعوبات عملية، منها ضعف الاستعداد المهني لدى عدد من المعلمين أو البيروقراطية السائدة في النظام التعليمي، مما قد يقلل من إمكانية استمرار الفكرة طويلة الأمد. تشير هذه النتائج إلى وجود دلالة قوية على أن تطوير مهارات التفكير النقدي يقوّي اكتساب اللغة الثانية، وهو ما يتماشى مع أفكار فيغوتسكي حول التعلُّم الجماعي. لذلك، تحتاج كليات وجامعات حائل إلى التركيز أكثر على أساليب التعليم الهجين. عوضًا عن الاعتماد فقط على الطرق القديمة، كما يؤدي الدمج بين المواد الدراسيّة إلى إيجاد بيئة تعليمية فعّالة، مما سهم بشكل مباشر في تحقيق أهداف رؤية 2030.

الكلمات المفتاحية: مهارات التفكير، التفكير النقدي، البيئة التعليمية، التعليم الهجين، الاستعداد المهني.

* أستاذ اللغويات التطبيقية المساعد، قسم اللغة الإنجليزية، الآداب والفنون، جامعة حائل، المملكة العربية السعودية.

للاقتباس: عبدالرحيم، ش. ج. م. (2026). التفكير النقدي والتطوير الشامل لمهارات اللغة الإنجليزية في جامعة حائل: تعزيز كفايات رؤية 2030، *الآداب للدراسات اللغوية والأدبية*، *8*(1): 586- 602 https://doi.org/10.53286/15rx7402

1. Introduction

The demands of the 21st-century global economy have fundamentally reshaped the required competencies for university graduates. No longer is specialized knowledge sufficient; rather, success hinges upon a graduate's capacity for adaptability, complex problem-solving, and continuous learning (Li, 2020). These capabilities are rooted in two essential skill sets: advanced English language proficiency for global communication and access to knowledge, and robust Critical Thinking (CT) skills for analyzing and evaluating information.

1.1 Global Context

Globally, English remains the dominant lingua franca of science, technology, and commerce, positioning it as an indispensable tool for accessing and contributing to the knowledge economy. Concurrently, CT is recognized by international bodies and educational experts as a prerequisite for navigating the complexities of modern civic life and workforce challenges (Facione, 2015). The effective integration of these two domains has become a critical pedagogical imperative for higher education institutions worldwide, demanding curricula that treat language and thought as intertwined processes.

In the Kingdom of Saudi Arabia, this imperative is amplified by the Saudi Vision 2030 strategic plan. The Vision explicitly mandates a radical overhaul of the education system to produce a knowledge-driven, innovation-capable workforce that can lead economic diversification in sectors like renewable energy, tourism, and finance. This transformation requires institutions, particularly regional anchors like the University of Hail (UoH), to align their curricula with national goals by cultivating graduates who can think critically *in* English to tackle complex, real-world problems. Despite this clear directive, the challenge remains in designing and empirically validating a pedagogical model that effectively achieves this dual outcome within the existing institutional constraints.

1.2 Problem Statement and Research Gap

Traditional English Language Teaching (ELT) in many regional contexts, including Saudi Arabia, has historically prioritized discrete skills and rote memorization, often due to high-stakes, grammar-focused assessments (Alghamdi, 2018). This approach leads to a functional proficiency that often falters when faced with advanced academic and critical tasks such as synthesizing research or evaluating opposing viewpoints. This conventional methodology treats language acquisition and cognitive development as separate, sequential processes, neglecting the synergistic relationship where linguistic fluency supports higher-order reasoning, and vice-versa (Vygotsky, 1978; Wallace, 2003).

While the theoretical necessity of integrating CT with Holistic English Language Development (HELD) is well-established in the literature, there remains a significant empirical research gap. Specifically, there is a

pronounced lack of quantitative, contextually validated studies demonstrating the measurable impact and effectiveness of a formal, integrated CT-HELD curriculum within the unique socio-cultural and educational environment of a Saudi Arabian university. Most research in the region is descriptive or survey-based.(Al-Segh,2021) yet Furthermore, the existing literature offers minimal insight into the *qualitative challenges*—specifically faculty preparedness, student perceptions, and institutional barriers—that accompany the implementation of such a radical pedagogical shift, all of which are crucial for scaling up a successful model.

1.3 Study Aims and Significance

To address this empirical and methodological gap, this study employs a mixed-methods, quasi-experimental design to investigate the efficacy of a CT-HELD integrated curriculum at the University of Hail. The primary aims are twofold:

1. To quantitatively measure the impact of the CT-HELD curriculum on student gains in both standardized Critical Thinking skills (CCTST) and Academic English Proficiency (AEPT).
2. To qualitatively explore the experiences of students and faculty to identify the key mechanisms for success and the primary barriers to sustained implementation, providing actionable data for policy change.

This research carries significant theoretical and practical significance. Theoretically, the study provides robust empirical data to validate the synergistic model of CT and L2 acquisition, particularly in an under-researched Middle Eastern EFL context. Practically and politically, the findings offer a proven, evidence-based pedagogical roadmap for UoH and other Saudi higher education institutions to directly operationalize the educational mandates of Saudi Vision 2030, ensuring the production of graduates with the advanced competencies necessary for national economic transformation.

Given how CT connects with HELD in theory, and because the goals of the research point in this direction, one central idea was put to the test

Those getting the combined CT-HELD approach tend to show stronger gains in critical thinking - alongside better command of academic English - when placed beside peers following standard lessons. Progress appears more pronounced where teaching blends these methods rather than keeps them separate.

When critical thinking grows stronger in the experimental group, English skills tend to rise at the same time. This pattern fits a combined approach to learning. Results suggest both abilities develop together. The link appears consistent across participants. Such alignment points toward shared growth mechanisms.

2. Literature Review

2.1 Defining Critical Thinking (CT)

Thinking critically means carefully processing thoughts to assess and combine ideas that shape choices and beliefs (Ennis, 2011). A range of abilities supports this - breaking down facts, drawing conclusions, making judgments, clarifying reasoning, and adjusting one's own thinking (Facione, 2015). Within education, it shifts students away from simply absorbing information, nudging them into deeper involvement, asking questions, forming sound opinions.

2.1 The Synergistic Relationship between Critical Thinking (CT) and Holistic English Language Development (HELD)

The mix of Critical Thinking (CT) and full-range English learning works best when each part helps the other move forward. Instead of working alone, they feed into one another naturally over time. CT means slowing down to think clearly about what makes sense or what step to take next (Ennis, 2011). It's made up of mental skills - some highlighted by Facione (2015) - like sorting ideas apart, making logical guesses from clues, judging whether something sounds true or strong, along

needs a solid grasp (HELD), along with sharp analysis (CT).

HELD moves teaching from just grammar drills to building listening, reading, speaking, so writing at once using tough mental work. It works because learning sticks better when you use words to explore real ideas - instead of treating them like a test goal. That's way different from old-school programs focused on right answers more than saying complex things well.

This mix of methods ties back to Vygotsky's 1978 idea called Sociocultural Theory. Instead of just learning passively, speech acts work like tools in the mind, guiding deeper thinking. When learners use English to dive into tough topics - like judging moral issues, breaking down research papers, or arguing different sides - they're working inside their Zone of Proximal Development. That zone? It's where someone manages a challenge only when helped, not alone. In this learning path, putting tough ideas into words using a second language builds speaking skill and sharp thinking at once - because shaping an opinion in another tongue means working with tricky grammar bits like soft modals or give-in phrases. Take making a strong case: it pushes you to use exact tools that match how deeply you're reasoning. (Wallace, 2007) pointed out that reading carefully isn't separate from growing language - it's part of it, so lessons should mix brain work with word practice. This combo creates a cycle where better thinking helps tackle hard texts and build subtle points, while richer language gives the right terms and flow to shape those thoughts clearly.

2.2 Pedagogical Imperatives for Integration

Achieve growth in CT plus HELD means moving beyond old-school teaching that focuses on repetition. Instead, classes should revolve around learners - pushing them to apply language when thinking things through or tackling challenges - so what they study matches actual job needs. The integration model builds on interactive, meaning-driven methods - especially Content-Based Instruction (CBI), but also Task-Based Language Teaching (TBLT) (Nunan, 2004). Instead of dry grammar practice, CBI uses engaging real-life topics - say, climate policies or world markets - to teach language skills. Learners must process challenging ideas using English, so comprehension goes deeper. By tying language use to actual context, it encourages students to move beyond memorization into genuine critical thinking. With TBLT, learners tackle substantial activities - such as drafting reports or delivering arguments - not just repetitive drills. These tasks demand reasoning and communication, which pulls students into active language use. Solving realistic challenges becomes the main way to grow critical thinking, because success depends on mastering the necessary vocabulary and structures. Getting students to think more deeply in English classes means using clear teaching methods on purpose. A focus on learners turns lessons into active mental workouts. Discussions with support push pupils to look at different views, weigh facts carefully, while forming logical points quickly - this needs strong language control to explain tough concepts well. Writing that asks for personal insight leads to greater self-awareness during study, along with sharper judgment when reviewing material, building both analysis skills and precise writing for detailed debates. When these tasks become part of regular work, teaching connects careful reasoning directly to gaining advanced language ability. Socratic talk: by asking thoughtful questions one after another, teachers help learners rethink what they take for granted - sharpening how they reason while guiding them to grasp ideas more deeply. This method mirrors real curiosity in action; at the same time, it builds skill in paying attention and replying well during conversation. Even though the benefits sound good in theory, putting them into practice in Saudi Arabia runs into real problems. One big issue? The strong tradition of learning by repetition sticks around, thanks to exams that reward memory work more than understanding (Alghamdi, 2018). But here's what research keeps pointing to (Habbsh, 2022): teachers aren't fully prepared. Many were trained using old-school techniques - like grammar drills or step-by-step skill teaching - and haven't learned how to teach critical thinking, let alone grade it properly. They need targeted support: creating realistic assignments, leading thoughtful class talks, plus building clear scoring guides that judge logic and thought process - not just spelling or grammar slips.2.3 Strategic Alignment with Saudi Vision 2030

This research answers Saudi Arabia's Vision 2030 plan in a focused way. Instead of relying on memorization, education must now build thinkers who adapt, solve issues, or tackle tough situations. The

goal? To create an open economy run by skilled minds ready for worldwide competition. Knowing English opens doors - to teamwork across borders, access to tech advances, and participation in science. Critical thinking drives how well people actually use that language. So this CT-HELD approach isn't just about speaking - it shapes workers who bring smart insights to new industries and services growing across the country.

The University of Hail helps push forward the plan for fair development across regions. Situated in Hail Province, it trains many locals so they can thrive in a modern economy focused on smart, high-skill jobs. Using a well-checked CT-HELD program, the university matches its courses to national and international levels. This move backs two main aims: improving learning through globally aligned teaching and research, while also supporting fairness by equipping Hail's youth for emerging job markets.

Even though policies say it's needed, there's still little real-world data on how well a structured CT-HELD program works in Saudi universities, given their unique cultural and academic settings. This research aims to close that gap by testing whether the program actually helps - using hard evidence - and also looking into what gets in the way during rollout, so educators can make smarter changes based on actual results.

3. Methodology

3.1 Research Design and Participants

Research Design

This study used both numbers and personal insights to look at results (Creswell, 2014). It worked well because it could show how the CT-HELD program affected thinking skills and language learning, yet also included real-life stories that helped make sense of those numbers. On one hand, there were tests before and after teaching; two separate groups took part - the first got the new combined course, whereas the second stuck with regular English lessons. Since students weren't shuffled randomly into classes - but stayed in their usual ones - it counted as a near experiment, mainly due to school rules and practical limits.

Participants and Setting

The research took place at UoH in Saudi Arabia, looking into learners from the English Department.

A hundred male undergrads in their second year took part. One group of fifty tried the new method, while the other fifty kept doing things the usual way. They all started with about the same score on an English test before anything began. This made sure both sides had roughly equal language skills at first.

Twelve more people took part in the deeper look at what happened after numbers were gathered. These individuals stood apart from the main set of 100 learners tested earlier. Among them: eight came from the test group - chosen carefully because they scored either high or low when progress was measured numerically. Four others were teachers who taught the course material. By mixing these types of voices on purpose, the conversations could reflect different real-life viewpoints tied closely to why certain patterns showed up in data. That opened space for clearer sense-making around how the teaching approach worked - along with what got in its way during delivery

3.2 The CT-HELD Intervention

The program lasted 15 weeks - one whole school term - with each group getting six hours of class weekly, which adds up to 90 hours altogether.

Control Group (CG) Curriculum

The CG got the usual English course built around basic skills. Instead, classes zeroed in on clear grammar points, word lists, plus understanding texts using separate tasks taken out of real context. They stuck closely to the set book format and testing style but didn't include methods for deeper thought or subject-based breakdown.

Experimental Group (EG) Curriculum

The EG got the CT-HELD curriculum - critical thinking built right into every language activity. It was shaped around Task-Based Language Teaching, along with Content-Based Instruction, focusing on tough real-world themes tied to Saudi Vision 2030; like how to keep local environments stable or handle shifts in the economy.

Key features of the CT-HELD curriculum included:

Lessons set aside time to focus on key thinking skills from Facione (2015), like breaking things down, drawing conclusions, or judging arguments - so learners examined a passage to spot hidden beliefs before checking how strong the proof was.

Teachers asked thought-provoking questions step by step - this pushed students to rethink their ideas while discussing texts or doing reading work, showing them how deeper thinking sounds out loud.

Students tackled assignments combining high-level language skills with thinking deeply - like crafting persuasive essays using outside sources, while also joining organized debates tracked for progress.

The curriculum brought in simple problem-solving ideas - like breaking tasks down or spotting patterns - not for programming, but to help organize thoughts when writing or talking. Instead of jumping straight into an essay, kids learned to split tough prompts into smaller parts first. This made their answers clearer and more organized. By pulling big questions apart early on, they built stronger arguments step by step. It's like planning the pieces before putting them together. That way, logic flows better from start to finish.

3.3 Instruments and Data Collection

Quantitative Instruments

1. California Critical Thinking Skills Test (CCTST):

Goal: check real progress in basic thinking skills.

The CCTST (Form B) is a widely used test in colleges around the world. Yet, it measures five core thinking abilities - like breaking down info or drawing conclusions. While built from 34 questions in multiple-choice format, each one leans on real-life situations. Instead of just facts, it checks how well someone solves problems they might face. So, responses show actual reasoning skill, not memorization.

Step: Everyone took the CCTST before starting - then again at the end, after the program finished.

2. Academic English Proficiency Test (AEPT):

Goal: Check how well someone uses high-level English in real-life ways.

The AEPT's an internal test that's been checked for accuracy - it judges skills tied to doing well in school. Instead of separate parts, it mixes tasks like understanding college-style talks with tough reading that needs connecting ideas. One section asks you to write a structured argument using clear logic and varied sentence forms. This part checks how well you build essays, use detailed grammar, yet keep thoughts flowing naturally together.

Test steps: AEPT given before and after training. Essay answers graded without knowing who wrote them - two outside graders did scoring separately. They used a clear breakdown method, checking argument layout plus complex sentence use. This helped make sure ratings matched closely between scorers.

Qualitative Instrument

Semi-Structured Interviews: We put together a clear plan to dig into what people really felt. These talks had loose guidelines - so we could go off script when needed - to get beneath the surface, digging deeper than numbers alone by asking follow-up questions that uncovered real stories behind the data, zeroing in on:

Learners: How involved they seem, how confident they feel, struggles with the updated lessons, also how clearly they grasp the connection between computational thinking and speaking English well.

Faculty: How ready they feel teaching CT, what hurdles pop up during rollout, whether past training worked well, also their take on how well the course fits broader educational aims.

3.4 Data Analysis Procedures

Quantitative Data Analysis

All numerical information was examined through SPSS software, specifically release 26. Different statistical tools helped process the figures accurately. The analysis relied on this program due to its consistent performance. Researchers used it to ensure reliable outcomes from collected measurements.

1. Descriptive statistics included averages, spread measures, also asymmetry values - these applied to both initial and follow-up test results. Various indicators helped summarize score patterns across testing phases instead of relying on single metrics alone.

2. Baseline similarity: Pre-test results were compared using t-tests to check if the EG and CG started at similar levels on both the CCTST and the AEPT.

3. Hypothesis Testing: For group comparisons, adjusting for small starting differences, an ANCOVA was applied to post-test results - using pre-test scores as a covariate. Instead of simple comparison, this method accounts for initial variation across participants. The analysis focused on adjusted outcomes, ensuring fairer contrasts between groups. By including prior performance, it reduced bias from unequal baselines. This approach improves accuracy when evaluating treatment effects. Rather than ignoring early data, it integrates them into the main model.

4. Effect size: For every meaningful result, Cohen's d was used to assess how strong the effect of the intervention really was - showing its real-world relevance through standardized differences.

5. A correlational analysis used Pearson's r to measure links between CT and HELD gain scores, examining whether the combined curriculum shows interactive effects. The aim was to assess if changes in one variable corresponded with shifts in another, using data from both domains to explore potential alignment. Rather than assuming direct causation, this method focused on pattern recognition across results. Each score change was compared alongside its counterpart to detect consistent trends. This approach allowed insight into how improvements in critical thinking related to gains in higher-order learning dispositions.

Qualitative Data Analysis

Transcripts from interviews got examined through thematic analysis (Braun & Clarke, 2006).

1. All interviews got written down exactly as spoken. After going through each transcript several times, basic labels were created to highlight key statements or expressions.

2. Patterns emerged when codes were grouped into wider themes. These themes showed consistent repetition across data. Grouping helped identify common ideas without overlap. Connections between concepts became clearer through this process.

3. Themes like Enhanced Engagement or Faculty Readiness Gap were checked, adjusted, while being made more precise - so each one fits the data well; also helping clarify the numerical findings. Instead of just listing ideas, connections between patterns and outcomes were stressed through careful revision. Definitions became sharper over time due to repeated comparison with evidence. This process ensured relevance without losing detail. Each theme was tested against responses to confirm consistency.

3.5 Ethical Considerations

The research followed ethics rules set by the University of Hail's IRB. While signing up, each participant gave written consent and received clear information about privacy protection. Although involvement was optional, everyone knew they could leave later if needed. Even after collecting responses, personal details were removed; files went into a protected system where only researchers had access.

4. Results

4.1 Quantitative Findings

The examination of early and late test results supported the main idea: using combined CT-HELD teaching led to clearer gains in critical thinking abilities as well as broader language progress when contrasted with standard methods.

4.1.1 Baseline Equivalence

T-tests comparing pre-scores from the CCTST and AEPT revealed no meaningful gap between the Experimental Group and Control Group. Because results were similar at baseline, later changes in post-test outcomes can likely be linked to the treatment applied.

Measure	Group	N	Mean (SD) - Pre-test	t	p
CCTST	EG	50	14.5 (2.1)	0.81	0.421
CCTST	CG	50	14.2 (2.4)		
AEPT	EG	50	62.8 (4.5)	1.15	0.252
AEPT	CG	50	61.9 (4.8)		

4.1.2 Critical Thinking (CCTST) Gains

A single ANCOVA was run on the CCTST post-test results, with pre-test scores acting as a control variable. This showed that the treatment had a significant impact.

The findings showed a clear gap between the two groups, with the Experimental Group performing better ($F(1, 97) = 28.95$, $p < 0.001$). While the EG had an average post-test result of 19.1 (SD = 2.0), the CG scored lower at 15.8 (SD = 2.2).

The effect size - computed with Cohen's d - came out to $d = 1.49$, which counts as very large. Such a value indicates that the CT-HELD approach didn't merely yield a significant statistical result; instead, it brought about a practical boost in student reasoning. Notably, the experimental group showed strongest

progress in Inference and Evaluation on the CCTST, areas specifically addressed through Socratic questioning and argument-based exercises.

4.1.3 Holistic English Language Development (AEPT) Gains

A further ANCOVA examining AEPT post-test outcomes revealed a clear advantage for the Experimental Group ($F(1, 97) = 15.62$, $p < 0.001$). While the EG had an adjusted average of 74.3 (SD = 5.1), the CG scored lower at 66.8 (SD = 4.9).

The effect size for HELD skill was $d = 0.89$, which counts as large. Although this improvement is notable, embedding computational thinking didn't harm language progress - instead, challenging, meaning-based activities may have boosted performance by pushing learners to apply diverse, intricate sentence patterns when expressing analytical ideas. Learners in the experimental group showed stronger command of scholarly writing, marked by clearer structure, improved linking of concepts, and more precise transitions.

4.1.4 The Synergistic Correlation

A closer look at how better critical thinking ties to progress in academic English comes from a Pearson correlation test. Post-test results minus pre-test ones gave gain values for each student in the group studied - fifty learners total. These numbers came from two tools: one measured reasoning skills, and the other checked grasp of scholarly language. Tracking only shifts after teaching began helped reduce bias from earlier skill differences. What changed during the course matters most here.

Looking closer, better critical thinking is tied closely to stronger English growth, $r(48) = .68$, $p < .001$. Students climbing the highest in one area usually rose just as much in the other. Data like this fits well with the idea that thinking deeply and using academic English feed into each other. When learners work through tough ideas while shaping arguments in English, each skill lifts the other - especially when taught together.

4.2 Qualitative Findings: Thematic Analysis

The analysis of semi-structured interviews revealed three main patterns that helped explain the numerical findings: first, greater involvement and confidence; second, uneven preparedness among instructors; third, structural obstacles affecting execution.

4.2.1 Theme 1: Enhanced Engagement and Self-Assurance

Students in the Experimental Group mostly said they felt more driven, while also showing stronger interest in what was taught - quite different from how routine grammar classes are usually seen. Discussing real issues like local environmental efforts or young people's part in Saudi Arabia's future plan helped them connect faster with the language. Instead of seeing words as abstract rules, many began using them actively because topics mattered to their lives right now.

Talk from students in open-ended talks backed up the idea of stronger involvement and personal confidence. Not long ago, one learner put it clearly - S7 said they used to fixate on proper grammar alone. Now things have changed; expressing thoughts feels natural, even necessary. The weight shifted - from perfect sentences to what actually gets communicated. Another student, S4, brought up how talking about real-life topics tied to their path ahead lit a spark to speak up more in English. Ideas started flowing because the subject matter felt close to home. Then came S12's take: speaking up in class now carries worth, like being heard truly matters. Each story carried a similar thread - noticing value in exchange, not just error-free speech. Learning stopped revolving around rules and began circling meaning. Confidence grew where relevance lived. Classroom dynamics tilted toward dialogue instead of drills. Moments of hesitation gave way to moments of contribution. Behind each comment sat a quiet transformation - one measured earlier through numbers but now given shape through words.

4.2.2 Theme 2: The Faculty Readiness Gap

Folks teaching the class said they liked the idea of mixing skills, yet plenty admitted they did not feel ready to handle the parts focused on reasoning - especially how to grade them. Grading became a main sticking point. One teacher put it plainly: "Checking grammar has rules, but judging whether an argument makes sense? That seems up for interpretation. We got just one training session. What would help more is having real grading tools plus someone to talk through examples with over time." Another educator agreed, adding, "The concept sounds good in meetings, but still turning that into scores each student can trust week after week? That part takes way more support than we have." Taken together, these remarks point toward something clear - teachers need steady guidance rooted in actual classroom tasks if they're going to grow comfortable shaping lessons and judgments around deeper thinking.".

4.2.3 Theme 3: Institutional Barriers to Implementation

A few teachers said putting together blended lessons took too much effort, especially when daily duties already fill their days. One explained how planning deep classroom talks felt pointless if tests only checked small grammar rules later on. Students noticed it too - what they practiced often did not match what got measured. When grading stays fixed on old formats, trying new ways feels risky. The gap between fresh teaching styles and long-standing exams makes lasting change hard to keep going.

5.1 Alignment with Recent Empirical Research

What we found matches up well with earlier work around the world on blending language teaching with clear thinking skills. Instead of just adding ideas together, some recent projects build reasoning tasks right into lessons - and see strong results. One such project showed students learning English improved their ability to analyze when they worked through activities asking them to judge and combine information. That lines up

closely with what happened here: big steps forward in how learners handled inference and judgment. Notably, those jumps were clearly seen in test scores showing deeper engagement over time.

A jump in academic English skills - sizeable, clear - fits with the idea that real thinking effort in language builds ability better than drills on grammar alone. Work by Bryant and Allaf (2022) lines up here: students doing projects in English wrote with more depth and flow than others not in such tasks. That pattern shows again in our group, where essays grew tighter, more layered in how they were built.

Above all, a clear link (r = .68) between critical thinking and language growth backs up the connected learning idea suggested long ago by Vygotsky (1978) and later expanded by Wallace (2003). What stands out here is how relevant this becomes within Saudi classrooms, where solid data on teaching methods tied to these skills has been hard to find, as Al-Seghayer (2021) pointed out just recently. Findings like ours do more than echo earlier global patterns - they bring fresh, local weight to combining thought and language in education.

Still, what we found fits into bigger problems already talked about. Our look at how prepared faculty really are matches up with Alghamdi's 2018 take on deep-rooted obstacles in Saudi colleges teaching critical thinking. On top of that, mismatched assessments - what we saw as a roadblock - echoes Habbsh's (2022) worry over new teaching styles clashing with old grading ways. So even though combining methods works well, making it work widely means tackling those structural hurdles, something many number-focused reports tend to skip.

5.1 Interpretation of Quantitative Synergy

The notably high effect size (d=1.49) for CT improvements in the Experimental Group stands out clearly. Such a result implies the CT-HELD approach went beyond slight academic gains, leading to deep shifts in key thinking skills - especially reasoning and judgment. In parallel, the substantial impact on HELD mastery (d=0.89) shows focusing on critical thinking didn't weaken language progress; rather, it seemed to accelerate it.

This dual boost clearly backs Vygotsky's Sociocultural Theory from 1978. A solid link ($r = 0.68$, $p < 0.001$) between improvements in CCTST and AEPT results shows reasoning in English pushes growth in language complexity and smoothness. Because students had to build structured, fact-supported claims (CT) inside the combined course, they ended up mastering advanced grammar and linking tools (HELD), which fueled ongoing improvement. These outcomes question old teaching methods that split thinking and language use, instead supporting their necessary integration during lessons.

5.2 The Mechanisms of Change

The success of the EG came from using Task-Based Language Teaching alongside Content-Based Instruction. Using material tied to Saudi Vision 2030 goals made lessons more meaningful; this helped push learning into students' Zone of Proximal Development. Activities like analyzing policies or engaging in Socratic discussions redirected attention away from memorization toward practical language use for thinking tasks. What set the EG apart was this demand to apply language meaningfully - unlike the CG, which showed limited growth under traditional drill-based methods. Student feedback on higher engagement and confidence suggests motivation grew through relevant topics plus active application, supporting their stronger outcomes.

5.3 Contextualizing Challenges and Sustainability

The results from interviews help explain why numbers alone don't tell the full story at the University of Hail (UoH). Instead, two key issues - Faculty Readiness Gap and Institutional Barriers - pose serious risks to keeping the CT-HELD approach going over time. Although the method works well in practice, staff expressed confusion about evaluating critical thinking skills reliably; this issue has come up before in similar settings (Alghamdi, 2018). Because of tight testing demands focused on single correct answers, teachers feel pulled away from teaching deeper cognitive processes. Without targeted training efforts and changes to end-of-course exams, expanding the project across departments may fail.

5.4 Strategic Implications for Saudi Vision 2030

The success of the CT-HELD approach matches closely with Saudi Vision 2030's aim to build a capable, learning-driven labor force. Rather than just supporting skill development in theory, this research offers real-world proof of a teaching method that consistently delivers graduates fluent in both high-level English and strong analytical reasoning - skills vital for international engagement and tackling intricate challenges. Since the model has been tested and confirmed effective, the University of Hail can shift focus from broad strategic goals toward concrete updates in course design, enhancing its contribution to regional progress within national change efforts.

5.5 Limitations

This study has some clear limits. Although pre-test scores were adjusted using ANCOVA, the use of existing classes - due to practical barriers - means selection effects might still play a role. Instead, focusing on one university and only male students narrows how widely the results can apply, pointing toward testing similar ideas at other schools and with women later on. Still, because the program lasted just one semester, longer follow-ups are needed to see if gains last over time.

6. Conclusion and Recommendations

6.1 Summary of Findings

This research showed that combining Critical Thinking teaching with broad-based English learning leads to better results than standard methods. While the test group improved noticeably in thinking abilities (Cohen's $d = 1.49$), their overall language progress was also clear (Cohen's d = 0.89). Moreover, these two advancements were closely linked (r = 0.68). Although interviews confirmed higher student involvement, they also pointed to weak teacher preparation and tight program timelines.

6.2 Contribution to the Field

This research makes a dual contribution:

1. The study provides solid numbers supporting how CT links with learning a second language - especially in a Middle Eastern setting where little research exists.

2. A contextual roadmap offers UoH a tested, data-driven approach to curriculum change - this aligns closely with the learning goals of Saudi Vision 2030.

6.3 Recommendations and Future Research

Given the results, these suggestions are put forward for the University of Hail along with comparable local universities1. Mandivate full faculty development: start an ongoing, focused training initiative aimed at improving teaching methods and assessment skills for CT-HELD, addressing the known gap in readiness. The program should cover clear scoring tools to assess thinking processes as well as techniques like guided questioning to stimulate critical dialogue.

2. Change how students are assessed: current testing methods should better match CT-HELD's goals - not relying on one-time exams, but using practical tasks like portfolios or written arguments instead; these approaches value deep thinking and advanced language skills more effectively.

3. Carry out long-term research: upcoming projects could follow learners from the Test Group to assess how lasting these improvements are after they finish school - also, similar trials might be done with groups of women or at different vocational schools nearby.

Through targeted changes, the University of Hail could build a practical approach to developing flexible, analytical minds - key for meeting Saudi Vision 2030 goals. While focusing on reform, it may strengthen workforce readiness by nurturing problem-solving skills needed across evolving industries. As priorities shift, education strategies must adapt - not simply expand outdated frameworks. Because future challenges demand agility, learning environments should emphasize reasoning over rote knowledge. When institutions align with national aims, they contribute more meaningfully to both the economy and society.

References

Alghamdi, A. (2018). Challenges in teaching critical thinking skills in Saudi universities: A case study. *Journal of Higher Education, 5*(2), 11-25.

Al-Seghayer, K. (2021). The reality of EFL research in Saudi Arabia: A bibliometric and methodological analysis. *Journal of Language Teaching and Research, 12*(3), 89-102.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Bryant, A., & Allaf, C. (2022). Project-based learning and academic English development in EFL contexts. *International Journal of Applied Linguistics, 32*(3), 450-467.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ennis, R. H. (2011). What critical thinking means: A summary of key traits and skills linked to it. *Informal Logic, 31*(1), 1-27.

Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment.

Habbsh, A. S. (2022). The impact of technology integration on students' English language proficiency. *International Journal of Language Studies, 16*(1), 121-140.

Li, L. (2020). Developing critical thinking in language education: A critical review of literature. *Thinking Skills and Creativity*, (37), 100683.

Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.

Paul, R. W. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world.* Center for Critical Thinking and Moral Critique.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching* (3rd ed.). Cambridge University Press.

Saudi Vision 2030. (2016). *Kingdom of Saudi Arabia Vision 2030*. Council of Economic and Development Affairs.

University of Hail. (2020). *Strategic plan 2020-2025*. University of Hail Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wallace, C. (2003). *Critical reading in language education*. Palgrave Macmillan.