

RESEARCH TITLE

The Relationship Between Dialogic Interaction and Analytical Thinking Skills in English as a Fourth Language among High School Students in Buqata, Golan Heights

Rose Suliman Abu Saleh Farhat¹, Balkees Suliman Amasha¹

¹ PhD in Teaching and Learning, An-Najah National University, Nablus, Palestine.

HNSJ, 2026, 7(9); <https://doi.org/10.53796/hnsj79/2>

Received at 05/08/2026

Accepted at 12/08/2026

Published at 01/09/2026

Abstract

This study investigated the relationship between dialogic interaction and analytical thinking skills among high school students learning English as a fourth language in Buqata, Golan Heights. A quantitative correlational design was employed with a convenience sample of 70 students from Grades 10 and 11. Data were collected using two researcher-developed instruments: the Dialogic Interaction Scale, which assessed reciprocity, cumulation, supportiveness, and purposefulness during English lessons, and the Analytical Thinking Test, which measured interpretation, analysis, inference, and evaluation. The findings showed that students obtained a mean score of 71.00 (SD = 8.45) in dialogic interaction and 69.47 (SD = 7.62) in analytical thinking. Pearson's correlation analysis revealed a statistically significant, moderate-to-strong positive relationship between overall dialogic interaction and analytical thinking ($r = .512, p < .001$). Dialogic interaction statistically accounted for 26.2% of the variance in analytical thinking scores. Significant positive correlations were also found between dialogic interaction and all four analytical thinking subskills, with the strongest association observed for analysis ($r = .480, p < .001$), followed by interpretation ($r = .360, p = .002$), inference ($r = .304, p = .010$), and evaluation ($r = .273, p = .022$). These findings indicate that students who engage more effectively in reciprocal, cumulative, supportive, and purposeful classroom dialogue tend to demonstrate stronger analytical thinking skills. The study highlights the educational value of structured dialogic activities in multilingual English classrooms, particularly for developing analysis and interpretation. Nevertheless, given the correlational design and the use of a convenience sample from one school, the findings do not establish causality and should be generalized cautiously. Further experimental and longitudinal research is recommended.

Key Words: dialogic interaction; analytical thinking; English as a fourth language; multilingual education; classroom dialogue; high school students; Buqata.

العلاقة بين التفاعل الحواري ومهارات التفكير التحليلي في اللغة الإنجليزية كلغة رابعة لدى طلبة المرحلة الثانوية في بقعائا، الجولان

روز سليمان أبوصالح فرحات¹، بلقيس سليمان عماشة¹

¹ دكتوراه التعليم والتعلم، كلية الدراسات العليا، جامعة النجاح الوطنية، نابلس، فلسطين.

HNSJ, 2026, 7(9); <https://doi.org/10.53796/hnsj79/2>

تاريخ النشر: 2026/09/01م

تاريخ القبول: 2026/08/12م

تاريخ الاستقبال: 2026/08/05م

المستخلص

هدفت هذه الدراسة إلى تقصي العلاقة بين التفاعل الحواري ومهارات التفكير التحليلي لدى طلبة المرحلة الثانوية الذين يتعلمون اللغة الإنجليزية كلغة رابعة في بقعائا، الجولان. اعتمدت الدراسة المنهج الكمي الارتباطي، وطُبقت على عينة متيسرة مكونة من 70 طالبًا وطالبة من الصفين العاشر والحادي عشر. جُمعت البيانات باستخدام أداتين من إعداد الباحثين: مقياس التفاعل الحواري، الذي يقيس التبادلية والتراكمية والدعم والهادفية أثناء حصص اللغة الإنجليزية، واختبار التفكير التحليلي، الذي يقيس مهارات التفسير والتحليل والاستدلال والتقييم. أظهرت النتائج أن المتوسط الحسابي لدرجات التفاعل الحواري بلغ 71.00 بانحراف معياري قدره 8.45، بينما بلغ المتوسط الحسابي لدرجات التفكير التحليلي 69.47 بانحراف معياري قدره 7.62. وكشف تحليل ارتباط بيرسون عن وجود علاقة ارتباطية موجبة، متوسطة إلى قوية، ودالة إحصائيًا بين الدرجة الكلية للتفاعل الحواري والدرجة الكلية للتفكير التحليلي ($r = .512$)، ($p < .001$)، حيث فسّر التفاعل الحواري إحصائيًا ما نسبته 26.2% من التباين في درجات التفكير التحليلي. كما أظهرت النتائج وجود علاقات ارتباطية موجبة ودالة إحصائيًا بين التفاعل الحواري وجميع المهارات الفرعية للتفكير التحليلي؛ إذ كانت أقواها مع مهارة التحليل ($r = .480$)، ($p < .001$)، تلتها مهارة التفسير ($r = .360$)، ($p = .002$)، ثم الاستدلال ($r = .304$)، ($p = .010$)، وأخيرًا التقييم ($r = .273$)، ($p = .022$). وتشير هذه النتائج إلى أن الطلبة الذين ينخرطون بفاعلية أكبر في حوار صفي تبادلي وتراكمي وداعم وهادف يميلون إلى إظهار مهارات أقوى في التفكير التحليلي. وتبرز الدراسة القيمة التربوية للأنشطة الحوارية المنظمة في صفوف اللغة الإنجليزية متعددة اللغات، ولا سيما في تنمية مهارتي التحليل والتفسير. ومع ذلك، فإن طبيعة الدراسة الارتباطية واعتمادها على عينة متيسرة من مدرسة واحدة لا يسمحان بإثبات العلاقة السببية، ويستدعيان توخي الحذر عند تعميم النتائج. وتوصي الدراسة بإجراء المزيد من الدراسات التجريبية والطولية في هذا المجال.

الكلمات المفتاحية: التفاعل الحواري، التفكير التحليلي، اللغة الإنجليزية كلغة رابعة، التعليم متعدد اللغات، الحوار الصفّي، طلبة المرحلة الثانوية، بقعائا.

Chapter One: Introduction

1.1 Introduction

The purpose of this study was to examine the relationship between the dialogic interaction and analytical thinking skill of high school students learning English as a fourth language in the study site, namely Buqata, Golan Heights. The research design adopted for this study was correlational quantitative with 70 students selected as the subjects of the study from Grades 10 to 11 at a high school. Two instruments were designed for this study: a Dialogic Interaction Scale (DIS) to evaluate the quality and frequency of students' reciprocal, cumulative and exploratory talk in the context of English lessons and an Analytical Thinking Test (ATT) that was designed based on four sub-skills of Facione's (1990) concept of critical thinking which are interpretation, analysis, inference and evaluation.

The mean scores for the dialogic interaction were 71.00 SD 8.45 out of 100 and for the analytical thinking were 69.47 SD 7.62 out of 100, respectively. A Pearson correlation was conducted between the two total scores, and the results showed a moderate to strong positive correlation ($r = .512$, $p < .001$) that accounted for ~ 26% of the variance in scores between the two variables. In terms of sub-skill, dialogic interaction was most strongly associated with analysis ($r = .480$, $p < .001$) and interpretation ($r = .360$, $p = .002$) and slightly less strongly with inference ($r = .304$, $p = .010$) and evaluation ($r = .273$, $p = .022$). The findings of both null hypotheses of the study were not accepted for level of 0.05.

The findings are interpreted from a sociocultural perspective of Vygotsky, dialogism of Bakhtin, and exploratory talk of Mercer and Wegerif, which all emphasize that spoken interaction is not only a means of information sharing but is also a means of thinking. Overall, the pattern of findings is generally consistent with previous studies that found connections between classroom dialogue and reasoning gains (Mercer, 2008; Alexander, 2020); but it extends this line of study to a fourth-language context in which students access an L1 (colloquial Arabic), an L2 (Standard Arabic), and then an L3 (Hebrew) before reaching an L4 (English).

The findings of this study suggest that meaningfully dialogic interaction is correlated with these learners' analytical thinking, but cannot prove that dialogic interaction is correlated with improved thinking. It is recommended for English teachers, curriculum planners, and future researchers of the context of Buqata and other contexts that are multilingual.

This idea of talking to oneself in one's head is not new. From Vygotsky's point of view (1978), higher mental functions are first formed in joint activity, and then passed to the individual in silence. At the very least, Bakhtin (1986) stated that language itself is a reply to another language, a signal to another language and that meaning is created in interaction with the listener and the interlocutor and not passed from one mind to the other. After several decades, researchers in regular classrooms noticed these ideas, and wondered a more real question: Do students think better when teachers intentionally structure more of this reciprocal reasoning talk? That question has been very encouraging and systems such as Thinking Together (Dawes, Mercer, & Wegerif, 2000) and Collaborative Reasoning (Reznitskaya et al., 2009) were developed for this purpose.

What none of that evidence has done, so far, is look closely at a classroom where English is not the second language a student is adding to a single mother tongue, but the fourth language sitting on top of an already active repertoire of three others. That is the specific gap this study tries to fill, using a group of high school students in Buqata as its case.

This study takes that question directly and looks at whether the amount and quality of dialogic interaction students experience in their English lessons are related to how well they perform

on tasks measuring analytical thinking. Rather than testing an instructional intervention, the study measures both variables as they are now among a group of high school students in Buqata and analyzes the statistical relationship between them.

1.2 Problem of the study

A complaint of teachers of English in Buqata is that students will write the correct sentences in a worksheet, but struggle to explain why in their writing, to know how to weigh two sides of an argument, or to make a conclusion that is not explicitly stated in the text. Little opportunity for students to "think aloud," disagree productively, or develop an idea over multiple turns of discourse in lessons that are primarily based on individual seatwork and teacher-student question-and-answer exchanges. Activities, if included, are short and tightly controlled, and do not necessarily provide the sorts of longer, reasoning rich exchanges identified in the literature as a hallmark of developing analytical skills (Mercer, 2008; Alexander, 2020).

The situation is complicated by the fact that English is the fourth language these students are learning. Arabic and Hebrew already have the same limited classroom time and cognitive resources and English is the second language with the least real-world applications outside of school (Cenoz, 2013). Whether dialogic teaching approaches that have been found to enhance thinking skills in other, less linguistically layered contexts are associated with better analytical thinking among these students has not been studied in the past to the best of the researchers' knowledge. This is the gap in the current study.

There's also a much more immediate problem of this. Teachers who are personally convinced of the value of classroom dialogue often feel reluctant to give up class time for it, at least in an exam-driven culture where a fixed set of vocabulary and grammar can feel safer. Without evidence from the local community connecting dialogic interactions with the educational outcome (analytical thinking), that hesitation is hard to defend on an educational basis. This study aims to provide that kind of local, concrete evidence that any decision to invest classroom time in structured talk can rest on something stronger than intuition.

1.3 Research Questions

In general, this study is designed to answer the following problem:

How is the relationship between dialogic interaction and analytical thinking skills in English as a fourth language in high school students in Buqata, Golan Heights?

This leads to four sub-questions:

- What is the level of dialogic interaction displayed by the high school students in English as a fourth language classes?
- What is the level of analytical thinking skills of the students in terms of interpretation, analysis, inference, and evaluation?
- Is there a statistically significant correlation between the students' total score on the dimension of dialogic interaction and their total score on the dimension of analytical thinking skills at $\alpha \leq 0.05$?
- Does the correlation between the two dimensions differ among the four analytical thinking sub-skills, and if so, which sub-skills are most closely related to the dimension of dialogic interaction?

1.4 Research Objectives

Based on the above questions, this study aims to:

- Describe the dialogic interaction of high school students learning English as a fourth language in Buqata.
- Describe the analytical thinking skills of the same group of students.

- Analyze if dialogic interaction is statistically significant to the overall analytical thinking scores.
- Analyze if dialogic interaction differs for each of the four analytical thinking sub-skills.

1.5 Research Hypothesis

The null hypotheses that are tested are as follows:

H₀₁: There is no statistically significant relationship, at level of significance $\alpha \leq 0.05$, between the total interaction scores and the total analytical thinking scores of High school students learning English as fourth language in Buqata.

H₀₂: At the $\alpha \leq 0.05$ level of significance, there is no statistically significant relationship between dialogic interaction scores and all four sub-skills of analytical thinking (interpretation, analysis, inference and evaluation).

1.6 Significance of the study

Theoretically, this study builds on existing research involving classroom dialogue and thinking (Mercer, 2008; Alexander, 2020; Wegerif, 2013), but introduces a new context where the class language is fourth language English and the students' repertoire includes many colloquial Arabic and Standard Arabic words along with Hebrew. The previous studies of dialogic teaching and thinking skills are mostly monolingual or bilingual, and it is not clear whether the same association occurs in the case of a third or fourth language, which this study addresses.

The practical implication is that, if, in this population, this dialogic interaction is found to be meaningfully connected to the analytical thinking, these English teachers can have a lever to pull that is low-cost and low-technology; they can just shift classroom time to accommodate structured, reciprocal talk rather than the addition of materials or equipment. Grade 10 and 11 are the years when national English exams affect students' educational futures, so any classroom practice that plausibly supports reasoning skills alongside language use is of direct, immediate interest.

Furthermore, there's a methodological contribution. Because no existing instrument had measured dialogic interaction and Facione-based analytical thinking sub-skills for fourth-language English learners at this age level, this study had to design its own tools. The Dialogic Interaction Scale and the Analytical Thinking Test, if they were refined further, could be useful to other researchers or teachers working with similar multilingual, additional-language populations and save them the same development work.

1.7 Definition of terms

Dialogic Interaction

Conceptual definition: dialogic interaction is classroom talk in which two or more speakers build on, challenge, and extend each other's work in a long-lasting and reciprocal manner instead of taking turns for some brief minutes. It is based on Alexander's (2020) account of dialogic teaching that talk should be collective, reciprocal, cumulative and purposeful and Mercer and Littleton's (2007) description of exploratory talk, in which reasoning is visible and open to challenge.

Operational Definition: In this study, dialogic interaction is the score of a student on the Dialogic Interaction Scale (DIS), which is a classroom observation test of the trained raters who observe students in a sample of English lessons that is a classroom observation instrument. The four rated dimensions reciprocity, cumulation, supportiveness, and purposefulness have a score of 25 out of 100 on dialogic interaction.

Analytical Thinking Skills

Conceptual definition: Analytical thinking, from Facione (1990) consensus definition of critical thinking, is the purposeful, self-regulated cognitive process of interpreting evidence, analysing arguments, drawing sound inferences, and evaluating claims to reach a well-founded judgment or conclusion.

Operational Definition: In this study, analytical thinking is the score a student receives on the Analytical Thinking Test (ATT), a paper-based test built on four sub-tests corresponding to interpretation, analysis, inference, and evaluation, each scored out of 25, for a total out of 100.

English as a Fourth Language.

To be precise, the term is employed to indicate the position of English in the language repertoire of students who have already learned, in some form, three other languages (colloquial Arabic, Standard Arabic and Hebrew) before beginning formal English teaching, and for whom English is the least commonly used of the four.

Operational definition: In this study the term applies to Grade 10 and 11 students at a comprehensive high school in Buqata who study English as a compulsory school subject within this more multilingual profile.

Analytical Thinking Sub-skills:

Interpretation, Analysis, Inference, and Evaluation. Conceptual Definition: These are the four skills of analytical thinking that are used to model the Analytical Thinking Test in our study, in line with Facione (1990, 2015). Interpretation is the knowledge and expression of a text or situation. Analysis is the identification of the relationships between statements, claims or evidence. Inference is the analysis of the reasonable conclusions and consequences that follow from given information. Evaluation refers to being able to judge the credibility of the claims as well as the logical strength of the arguments.

Operational Definition: In our study, each sub-skill is the score a student gets on the corresponding sub-test of the Analytical Thinking Test, scored out of 25 points.

Chapter Two: Theoretical framework

2.1 English as a Fourth Language

2.1.1 The Concept of Fourth Language Acquisition

Historically, language acquisition studies have concerned the learning of a first language (L1) and to a slightly lesser degree, a second language (L2). The study of third and subsequent languages is a relatively new area of study. It is Cenoz (2013) who refers to the acquisition of any language that is not the same as the one the learner speaks as the L1 and is acquired after one has already acquired at least one other language as the third language acquisition (TLA). The fourth language acquisition (L4A) is a step further of this framework, as it refers to the learning of a language by a person who has already mastered the use of the three previous languages - a scenario that has become very common in the multilingual societies around the world (Cenoz, 2013).

The difference between L1 and L2, L3 and L4 is not a question of counting languages only. Every new language that is added to the repertoire of the learner comes under different circumstances. L1 is learned by immersion in early childhood, without effort, and without being taught in a second language system. The L2 is normally learned later and usually with a certain amount of formal teaching and influenced by the structures of the L1 of the acquirer. The L3 and L4, which are learned even later, are under the pressure of all the previous languages and are exposed to a more complicated network of transfer, interference, and

resources competition than any of the previous languages (De Angelis, 2007, as cited in Cenoz, 2013).

In the case of students in Buqata, students learn English subsequent to the languages of the Arabic, Hebrew, and Aramaic already having found a cognitive and social presence in the students, leaving the acquisition of English language to its own set of circumstances that are not yet well delineated by either the traditional SLA theory or the research on L3 acquisition.

2.1.2 The Linguistic Profile of Students in Buqata

The students at the high school in Buqata grow up speaking colloquial Arabic at home and in daily social life. Standard Arabic is learned through literacy and formal or religious context. Hebrew is taught intensively from the early school years in order to be the language of the larger civic and administrative system of the school and is in use in various educational subjects. English comes second and is, for most students, a subject studied in school rather than what they do all day. The order of this pattern— colloquial Arabic first, Standard Arabic and Hebrew in conjunction with the language in the school years, and English as a fourth language for a classroom-bound language— sets Buqata apart from the bilingual and second language contexts that dominate the existing research literature on classroom talk and thinking skills.

This layered linguistic profile is important for the present study as it means that any relationship between dialogic interaction and analytical thinking in English lessons is happening in a context of great linguistic complexity where English is competing with three other languages for classroom time, attention and cognitive resources.

2.1.3 Multilingualism and Cognitive Development

A separate line of research asks whether using multiple languages changes the way people think generally, irrespective of what language is used at the time. Bialystok (2001), and later work with colleagues (Bialystok, Craik, Green, & Gollan, 2009) found that people who are working with more than one language are better at a series of tasks that require selective attention, ignoring irrelevant information, and changing between mental sets (e.g., skills that are usually under executive control). Cummins (1978) had reported a similar phenomenon years earlier. Children who are raised with more than one language are more likely to see language as an object in their own right, and to think about it in a more metalinguistic way.

These results do not automatically imply that Buqata's four-language students are somehow better analytical thinkers by default. What they do suggest is that the cognitive machinery these students bring to an English lesson, honed over years of dealing with multiple language systems at once, might make them particularly well placed to benefit from tasks that ask them to notice, compare, and reason about competing pieces of information provided the classroom actually gives them opportunities to do so out loud.

2.1.4 Teaching English in Multilingual, Fourth-Language Classrooms

Methods that are designed for monolingual or simple bilingual teaching are not always transferable to a classroom where four languages are already in play. Cummins (2017) argues that instruction works better when it works with a learner's full linguistic repertoire instead of trying to isolate the different languages into specific, walled-off places; that a student able to work with Arabic or Hebrew to provide some clarity on an English concept, for instance, is not cheating the system, but is making use of cognitive resources that exist. Cenoz (2013) also argues that multilingual learners often need more explicit support and automatic support in the new language because there is already so much attentional resources invested in the other three.

For a fourth-language English classroom in particular, this can be related to teaching methods that provide students with frequent and low-pressure opportunities to create spoken English about the real content and not the silent, individual seatwork approaches. Dialogic interaction, which is described in detail in Section 2.3, is one such approach, and it is one that this study is focused on.

2.1.5 Codeswitching and Translanguaging as Classroom Resources

A related line of research asks what happens when multilingual students are not so scared to go back and forth between their languages but are encouraged to use their language while trying for a new one. García and Wei (2014) use the term translanguaging to describe this kind of flexible, integrated use of a learner's entire linguistic repertoire and they argue that thinking of a student's languages as separate, non-interacting systems does not accurately describe how multilingual minds actually work in practice. For a Buqata student who is thinking about an English task in a more abstract manner and needs to refer to Arabic or Hebrew to clarify a term or to check something out is not an indication of a lack of English fluency but a natural and often useful feature of multilingual thinking.

This study did not manipulate codeswitching or measure it directly, and the Dialogic Interaction Scale scored talk in English. Nevertheless, the translanguaging literature is a good reminder that fourth-language students bring a wider and more flexible set of linguistic resources to a discussion task than a monolingual framework would suggest, which is consistent with the broader argument in Section 2.1.3 that multilingualism need not be treated as a deficit to work around.

2.2 Analytical Thinking Skills

2.2.1 The Concept of Critical and Analytical Thinking

Facione (1990) reported on the results of a Delphi study with a large panel of philosophers and education experts, which characterised critical thinking as a purposeful, self-regulated process of judgment where interpretation, analysis, evaluation and inference are given, along with an account of the evidential, conceptual and contextual factors in making the assessment. In this context, analytical thinking is not a separate faculty from critical thinking, but is one of the cognitive skills that are concerned with identifying the intended and actually meaningful relationships between statements, questions, or other representations used to convey beliefs and reasons (Facione 1990, 2015).

This study uses Facione's account because it breaks down a broad and at times vague concept into a set of skills that are observable and can be measured with reasonable consistency, and has already been used in language education research (see for instance Ma, 2023 on flipped classrooms and critical thinking).

Facione's account is not the only one of the most influential. Ennis (1985) proposed an earlier, similar framework for clarification, basic support, inference and strategy and tactics, which is very much in line with Facione's work even though the two frameworks were developed in different ways. The overlap between Ennis and Facione's accounts is clearly a reason to use either one: the skills we are describing are not just an artefact of the vocabulary of one research group. Halpern (1998) makes a different, but very much important point; critical thinking skills do not seem to transfer automatically from one subject or context to another. And more effectively when the knowledge is taught in terms of a clear understanding of the reasoning skills rather than in the subject content of the subject is done by focusing on the ability to think from one subject to another, that is how we are going to achieve more. This is very relevant to our study as it suggests that analytical thinking in English and dialogic situations is not always very well-suited to an analytical thinking class in mathematics or

science (and this is something we do not make this study look at as a part of our study but should still be highlighted for the research we do), and it should be flagged for future work.

2.2.2 Four Sub-skills: Interpretation, Analysis, Inference, and Evaluation

- Interpretation, Analysis, Inference and Evaluation. Interpretation is the ability to understand and express the meaning of a given piece of information, experience, or situation essentially, working out what a text, statement, or scenario is actually saying.
- Analysis is the ability to identify the intended and actual relationships among statements, arguments, or pieces of evidence used to express a belief, judgment, or opinion.
- Inference is the ability to identify and secure the elements needed to draw reasonable conclusions, to form hypotheses, and to consider the consequences that follow logically from a given set of facts or claims.
- Evaluation is the ability to assess the credibility of statements or other representations, and to judge the logical strength of the relationships among them.

Facione (2015) treats these as related but distinct skills, and studies that categorize them separately (as opposed to as a single composite score) often find that they do not follow the same path: someone might be good at understanding a passage but not good at making sound conclusions from it. This fact is why we have measured the four sub-skills separately in this study, rather than relying on a single combined score.

2.2.3 Analytical Thinking in Second and Additional Language Learners

Language and thought are not easily separated from one another, especially for learners who are learning in a language they have not fully mastered. A beginner who is trying to analyse an argument in English is doing two things at once - processing the language itself and thinking about its content. When language processing demands are high as are typically the case for a fourth language used mostly in the classroom there is less cognitive room for the reasoning task itself unless classroom practice has created enough automatic language use to free up that capacity (Cenoz, 2013). This is why dialogic practice, which gives students repeated, low-stakes opportunities to produce and process English while reasoning about real content, is of particular interest for fourth-language learners.

2.2.4 Assessing Analytical Thinking in Educational Research

Facione's framework has been the basis for a small family of assessment tools in recent years, with the most prominent being the California Critical Thinking Skills Test, which is based on the same interpretation-analysis-inference-evaluation model from Facione (Facione, 1990). Ma (2023) used a shorter, classroom-friendly questionnaire based on the same six-skill Facione model to measure gains after a flipped-classroom intervention that showed the framework can be tailored to a particular study population rather than the entire commercial instrument.

This is the same adapted approach adopted in this study. Rather than importing an existing test for fourth-language English learners at high school level, the researchers designed the Analytical Thinking Test for this group with Facione's four sub-skills as the organizing structure but also to generate content and vocabulary that would be appropriate for Grade 10 and 11 English classes in Buqata. The reasoning behind this choice and the methods used to evaluate the test's validity are outlined in detail in Chapter Three.

2.2.5 Analytical Thinking and Academic Achievement

Along with its impact on general life skills, analytical thinking is also fairly consistently associated with academic achievement. Ma (2023) found that improvements on Facione's thinking sub-skills post-more discussion-based teaching were linked to higher performance on

course assessments in the same terms, suggesting that the relationship between thinking skills and academic success is not just background information but something that students might be able to influence in teaching design. Halpern (1998) also makes a similar point about durability: thinking skills that are actively learnt and considered rather than picked up off the shelf will have longer-lasting academic impact.

In English language learning in particular, analytical thinking may be of double worth. It is certainly a skill that a Grade 10 or 11 student will always need to have to work well in writing and speaking English, and that is part of what they learn in English reasoning tasks in other subjects taught in a student's own language as well. This dual role is of a real-world relevance for the current study as we shall see in Chapter One when we come to the main point of this study.

2.3 Dialogical Storytelling

2.3.1 The Concept of Storytelling in Language Teaching

The storytelling task is among the oldest communicative activities in which people engage and one of the most natural contexts for the use of extended spoken language. The storytelling approach is applied in language pedagogy at all levels of education, including the various cultures in which it is employed, for purposes of vocabulary learning, reading comprehension, listening development and this study especially focuses on speaking.

The findings of Kirsch (2016) confirmed that sustained storytelling instruction in the first language classroom not only increased oral engagement in the language classroom, but also led students to use the language more meaningfully than structured exercises did, and to achieve vocabulary acquisition and communicative confidence which the latter failed to deliver. Avraamidou and Osborne (2009) held that narrative is a uniquely powerful form of communication because it structures information in a way that is natural and motivating to the human mind – beginning, middle, end – characters, events, meaning. They demonstrated that this narrative form allows learners to be motivated to write and listen to the language as a medium to convey meaning.

The function of storytelling in communicative language teaching is similar to what Richards (2006) calls fluency activity, which is an activity aimed at conveying meaning without paying attention to the form of language. Storytelling can place the speaking task in a story context, and thus shift the evaluative atmosphere away from more formal speaking tasks and provide opportunities for learners to generate longer, less stressful, and more spontaneous speech.

2.3.2 Exploratory Talk and Thinking Together

Mercer and his co-workers developed exploratory talk to describe a particular and especially productive form of dialogue in which speakers lay out relevant information, challenge and justify their views with reasons, then actively seek agreement before going on from there (Mercer & Littleton, 2007). Their Thinking Together program (Dawes, Mercer, & Wegerif, 2000) trained groups of school children to use ground rules for such dialogue and a series of classroom studies linked the training to concrete improvements in individual and group reasoning as well as to better pedagogical learning of science and mathematics (Mercer, 2008; Wegerif, Perez, Rojas-Drummond, Mercer, & Velez, 2005).

Wegerif (2013) later commented that human thinking is fundamentally dialogic in nature, as internal reasoning comes out of and still resembles the structure of dialogue with others, which leads to the theoretical idea that students who engage in more and better classroom dialogue should also have better individual reasoning skills like the kind of analytical ones studied in this study.

2.3.3 Dialogic Interaction in EFL and Additional Language Classrooms

Dialogic methods have been recently applied to language classrooms in a few studies that have been published with positive outcomes for spoken and thinking skills. Ma (2023) found that a flipped classroom design, with a shift in class time from lecture style presentation to discussion and more time spent in class, was associated with Facione's critical thinking skills improvement among university students. Studies of Socratic and dialogic questioning in EFL teacher training work show similar results, with more frequent dialogic engagement in language learners being associated with more reflective, analytically informed questions. Li (2016), reviewing research on the relationship between thinking skills and second language education, notes that language classrooms are a particularly good place to learn the skills of thinking, not least because they already require learners to put reasoning into words, yet they are very often not used to teach for short, closed exchanges.

2.3.4 The Teacher's Role in Dialogic Classrooms

None of this happens automatically simply by letting students talk more. Michaels, O'Connor, and Resnick (2008) outline a set of “accountable talk” strategies asking a student to say more, asking another student to restate what their peer has said, and demanding evidence in support of a claim that teachers use to keep discussion of the students’ ideas rigorous and inclusive. Left unsupported, student talk could just as easily be shallow or off task; it is the teacher’s moves that will determine if talk turns into the type of reasoning rich discussion that the theory in Section 2.4 describes.

Reznitskaya and Gregory (2013) put forth a more realistic theory for how dialogic teaching should change students in a more lasting way and to what extent so in particular: a more sophisticated understanding of knowledge itself (claims must be supported by reasons), better argument skills, and greater insight into the subject matter. Resnick, Michaels, and O'Connor (2010) recapitulate this argument under the phrase “talk builds the mind,” examining evidence that classroom talk, and not just the quantity of students’ talk, is predictive of better reasoning skills. In a study like this where dialogic interaction is measured by students' direct classroom observation instead of how much they talk, this distinction between quantity and quality of talk is a very important one.

2.3.5 Traditional Discourse Patterns Versus Dialogic Talk

It is important to specify the contrast between dialogic interaction and traditional discourse patterns. Sinclair and Coulthard (1975) were the first to describe the dominant pattern of classroom talk that is found in most conventional classes: a three part exchange in which the teacher begins with a question, the student answers, and the teacher evaluates the student's response before moving on; this is often referred to as IRF (Initiation Response Feedback). Mehan (1979) also noted a similar pattern, IRE (Initiation Response Evaluation), in his classroom observations, and these two accounts are widely used in the literature to describe teacher centred discourse even decades later.

The problem with the IRF or IRE pattern, in line with the theory of change proposed in Section 2.4, is not that it is bad teaching; it is that it structurally limits how much thinking a student should have to come up with. A student answering a single, closed teacher question and then receiving a teacher’s evaluation has little opportunity to build an extended argument, respond to a peer's differing view, or revise a position in light of a counter example all activities that Reznitskaya and Gregory (2013) treat as the actual mechanism by which dialogue is supposed to encourage thinking. This study was based on the Dialogic Interaction Scale in order to distinguish classrooms and student contributions that do not fit the IRF pattern from those that do.

2.4 Theoretical Foundations

The study is based on Vygotsky (1978) sociocultural theory of learning and in particular the view that higher mental functions such as reasoning can first appear in social and dialogic exchange in the lives of people before being internalised as individual, silent thoughts. Learning takes place most productively in the learner's Zone of Proximal Development this is the gap between what a learner can do alone and what they can do with the help of a more capable partner and dialogic classroom talk is one of the main channels through which this guidance is delivered (Lantolf, Thorne & Poehner, 2015).

2.4.1 Scaffolding and the More Capable Peer

Wood, Bruner and Ross (1976) gave Vygotsky's Zone of Proximal Development its most common practical translation and coined the term scaffolding to describe how a potentially more capable partner who might have been a tutor but can be a peer also takes on some part of the task that a learner cannot do at the moment, and gradually withdraws that support as the learner's skills build up. When the learner was working up to it, they identified some tutoring functions that enable them to do this: keeping the learner engaged in the task, streamlining it to a manageable level, keeping the learner focused on the goal, highlighting the things that matter and managing frustration, and bringing out a solution when needed.

In a dialogic English classroom, these scaffolding functions don't come from the teacher alone. A student who clarifies a peer's confused point more clearly or who models how to back up a claim with a reason is doing exactly what Wood and his colleagues described but from a different place, at the front of the room. This peer to peer model of scaffolding is one of the more concrete, moment to moment ways in which the greater theory of dialogic interaction and analytical thinking we have discussed throughout this chapter is believed to work in practice.

Bakhtin's (1986) dialogism offers something more specific at the level of utterance: meaning is something we build together in a conversation and not transmitted whole from one mind to another. Mercer's (2008) sociocultural discourse analysis brings these concepts into the classroom and makes them more concrete: how the quality of talk between teacher and student and students themselves determines exactly how we think and understand it and so on too.

Reznitskaya and Gregory's (2013) theory of change adds another layer of detail to the picture, suggesting that dialogic classrooms change students through three things: they learn that knowledge is created and held in place through argument rather than handed down by authority; that they practice and internalise the specific actions of argument (for example giving reasons, and reasoning about the other side); and that they understand the content of whatever the dialogue is about. On the English level in Buqata, this is: dialogic interaction ought to help us in analytical thinking not in some vague, general sense of "talking more," but by practicing the practice of saying what we think, having reasons for it, and responding to a different view (the activities are quite similar to Facione's interpretation, analysis, and evaluation skills).

All such perspectives have one expectation: students who regularly engage in reciprocal and cumulative reasoning discussion should be better analytical thinkers than those who do not engage in such frequent, closed, teacher driven talk in class. This is the expectation that we want to put to the test in this study.

2.4.2 Summary of the Theoretical Model

When the three theoretical strands are put together, the model behind this study is to roughly be described as follows. Vygotsky (1978) gives the broad argument that reasoning is developed in social interaction before it becomes thoughts in the public. Bakhtin (1986)

points out how talk can do that work: meaning in dialogue is developed collectively, not as a whole. Reznitskaya and Gregory (2013) narrow this down into a workable, testable mechanism: saying a position, giving reasons and responding to a different view is what builds analytical capacity, not some more diffuse benefit of talking by itself. Applied to English lessons in Buqata, the model of dialogic interaction (defined in terms of reciprocity, cumulation, supportiveness and purposefulness in Chapter Three) should be associated positively with analytical thinking, in terms of Facione's interpretation, analysis, inference and evaluation skills. This is exactly the prediction tested in Chapter Four.

2.5 The Relationship between Dialogic Interaction and Analytical Thinking

The connection between dialogic and thinking theory has already found some empirical support outside language classrooms. The Thinking Together programme has shown that children with structured, ground based group discussion demonstrated reasoning gains (Mercer, 2008; Wegerif et al., 2005) and most reviews of classroom dialogue research find some relationship between dialogue and reasoning or critical thinking (Alexander, 2020).

Reznitskaya et al.'s Collaborative Reasoning programme is perhaps the best evidence of this relationship. Researchers (2009) found that students who participated in structured, peer led discussions of open ended questions developed more sophisticated and more robust arguments, both in the discussion itself and later written works, and these students were taught in the manner of the traditional class discussion of the same texts. Kuhn (2016) makes a similar argument but from a slightly different angle, arguing that dialogic exchange gives students the opportunity to engage in argumentative reasoning (a claim against a counter claim) in a way that solitary reading or listening rarely would.

On a more general scale, the Education Endowment Foundation's national trial of dialogic teaching in England (Jay et al., 2017) showed that students on the basis of structured dialogic teaching had significantly greater improvements in English, mathematics, and science during a single school year (even if the intervention was based on the quality of classroom talk rather than subject content).

This is one of the most detailed and well controlled studies of dialogic teaching and is generally in line with the smaller scale findings in this section. Inside language education, the evidence is less tangible, but it is very much the same. Ma (2023) found statistically significant gains on Facione's thinking sub skills under a more discussion based teaching design. Studies on Socratic and dialogic methods in EFL teacher education systems also indicate an increase in reflective and analytical capability with ongoing dialogic engagement. However, none of these studies have been conducted with students learning English as a fourth language, and they are currently in competition to use the same classroom time and cognitive resources as three other languages. This works in such a setting that is previously unexamined.

2.5.1 Possible Moderating and Mediating Variables

The literature reviewed above points to a number of factors which could plausibly strengthen, weaken, or even explain a relation between dialogic interaction and analytical thinking, even though the correlational approach used in this study is not designed to test them directly. General academic ability is a clear candidate: a student who is more academically capable and speaks more than anyone else might be able to talk more and do better on a thinking test, and neither of these variables directly drives the other. Fluency in the three languages followed by English (colloquial Arabic, Standard Arabic, and Hebrew) is the second possibility in this context; if students who are strong in multilingual metalinguistic awareness (Bialystok, 2001; Cummins, 1978) can work in a fourth language, in that case they can both engage in a dialog with and explain in a way analytical about it.

The third option is that the relationship is mediated by something other than a language (such as confidence and a willingness to take verbal risks with language) through the ability to speak up, and therefore develop analytical skills and not dialogic interaction and analytical thinking are independent of each other: that makes it easy for one to think and analyze. The problem is defining these possibilities (confounding, moderation, and mediation in this context) is not so easy as one correlational study can do, but it is a natural next step to take and is not far from the experimental or longitudinal research we discuss in Chapter Five.

2.6 Previous studies

2.6.1 Studies on Dialogic Teaching and Thinking Skills

Mercer and Littleton (2007) reviewed a decade of classroom based research on dialogue and children's thinking and found that structured opportunities for exploratory talk were associated with better reasoning performance in the classroom through standardised reasoning tests before and after dialogic instruction. Wegerif and colleagues (2005) extended this line of research to other countries, in the United Kingdom and Mexico, and found similar reasoning gains in both settings and argued that the underlying mechanism, talk as a tool for thinking, did not depend on any specific cultural or linguistic setting.

Alexander (2020) summarised a wider international programme of classroom observation research and identified dialogic teaching as one of the few classroom interventions with evidence that was consistent to improve both spoken engagement and depth of reasoning, but he was careful to note that the effect differed depending on how well teachers implemented the core dialogic principles of reciprocity, cumulation and purposefulness.

2.6.2 Studies on Multilingual and Additional Language Learners' Thinking Skills

Bialystok and colleagues' (Bialystok 2001; Bialystok et al., 2009) research found that managing more than one language is associated with advantages in executive control tasks, such as selective attention and resistance to irrelevant information to make intelligent analytic decisions. Cummins (1978) also found similar advantages in metalinguistic awareness, the ability to consider language as a subject of reflection, among bilingual children against monolingual ones.

Ma (2023) is one of the few studies to link thinking skills measurement directly with a language adjacent instructional design, and found that university students with more discussion heavy teaching performed better compared to a comparison group on Facione's interpretation, analysis, inference and evaluation sub skills. This study did not use students working in more than one language nor a population comparable to the fourth language students in Buqata.

2.6.3 Studies on Classroom Talk and Reasoning Skills More Broadly

Besides the language education research we have already reviewed, there is a much larger body of general education research on this study's central question. The dialogic teaching trial conducted by The Education Endowment Foundation (Jay et al., 2017) for 76 primary schools and almost 5,000 pupils is among the largest controlled studies of dialogic teaching to date and the results suggest consistent, albeit modest, achievement gains on better classroom talk. Reznitskaya et al. (2009) showed that argument quality improved in elementary students after talking to teachers in a structured, peer led discussion group, and Reznitskaya and Gregory (2013) proposed a theoretical method in Section 2.4 through which educational attainment gains are believed to occur. Michaels et al. (2008) and Resnick et al. (2010) add a practical dimension and explain the teacher talk moves that seem to distinguish classrooms where dialogue is consistently supported in reasoning from those in which dialogue is not.

Taken together, the literature in this section is reasonably consistent, and yet almost none of it has been done in a language classroom, let alone in a fourth language classroom. Most of the studies cited above deal with native speaking students talking about the content in their first language. Whether the same mechanisms are in play when the medium of dialogue is itself a language that students are acquiring (and their fourth at that) is the question our study is designed to answer.

2.6.4 Research Gap

Three threads run through the literature reviewed above. Research on dialogic teaching consistently links structured classroom talk to improved reasoning. Research on multilingual cognition consistently links managing several languages to advantages in executive control and metalinguistic awareness. And a small but growing body of language education research links more dialogic, discussion based instruction to measurable gains on Facione's specific thinking sub skills. What's missing is a study that brings these three threads together in a fourth language setting: in one where dialogic interaction and analytical thinking are measured together, in an English classroom that is situated on top of an already existing three language repertoire.

To the best of the researchers' knowledge, no study has examined this relationship with such students as those in Buqata, where colloquial Arabic, Standard Arabic and Hebrew are already part of daily linguistic life before English lessons even start. This work fills that gap.

2.6.5 Summary of the Literature Review

This chapter has been about four basic topics: English being a fourth language in a multilingual setting; Facione's account of analytical thinking and its four sub skills; dialogic interaction and those mechanisms (from Vygotsky to Reznitskaya and Gregory) to link it to reasoning; and the current empirical evidence linking classroom dialogue and thinking skills in general and, more specifically, language classrooms. The common thread between all four areas is that reciprocal, cumulative and reasoning oriented talk (as opposed to short, closed and teacher driven) is associated with better thinking. However, until this study, it remains to be seen whether that is the case for students who speak English as a fourth language alongside colloquial Arabic, Standard Arabic and Hebrew.

2.7 Conceptual Framework of the Study

The conceptual framework for this study is to consider dialogic interaction as an independent variable and analytical thinking as a dependent variable and, based on the literature reviewed above, the two should be positively correlated. Dialogic interaction is divided into the four dimensions presented in the Dialogic Interaction Scale (reciprocity, cumulation, supportiveness and purposefulness) and analytical thinking is divided into Facione's four sub skills (interpretation, analysis, inference and evaluation). The framework does not provide a one to one relation between different dialogic dimensions and analytical sub skills; rather, it considers dialogic interaction as a single, combined construct that should be linked to analytical thinking as a whole and allows, as an empirical question rather than an assumed one, for the extent of this connection to be different from one to the next for each of the four sub skills. This is precisely the structure of the two research hypotheses of the study in Chapter 1 and the correlational analysis in Chapter Four.

Chapter Three: Methodology

3.1 Research Design

This study used a quantitative correlational design. A correlational design is not an experimental or quasi experimental one, as there is no teaching intervention or comparison of treatment to control groups, but rather both variables of interest (dialogic interaction and analytical thinking) are measured as they naturally appear for students and the statistical relationship between the two sets of scores is examined (Creswell & Creswell, 2018). This design is well suited to the research questions of this study, because our aim is to describe the association between dialogic interaction and analytical thinking as they are now in Buqata classrooms and not to test whether one teaching method causes an increase or a decrease in the variables. A quick reminder of what a correlational design can and cannot show.

A strong relationship between dialogic interaction and analytical thinking suggests that the two variables are rising and falling together and would be the theoretical expectation of Chapter Two. It would not, on its own, prove that dialogic interaction is linked to analytical thinking, because a third factor – general academic ability, for instance, or a student's overall engagement with school – could plausibly affect both variables at once. This limitation is addressed again in Chapter Five.

3.2 Participants

The study sample was 70 students in Grades 10 and 11 at a comprehensive high school in Buqata, picked by convenience sampling from classes that already were following the standard English curriculum for their grade level. Both grades were included to make the sample larger and to enable the correlational analysis for dialogic interaction and analytical thinking scores. Students with a formally documented learning difficulty affecting language processing were excluded from the sample since the analytical thinking test in particular assumes a typical developmental profile.

The school itself is a comprehensive mixed grade school that serves the town of Buqata and offers English as a compulsory subject at all grade levels, with Grade 10 and 11 students typically receiving four to five English lessons per week as part of the usual timetable. English classes at this level are based on the same national curriculum framework used throughout the wider school system and cover reading, writing, listening, and speaking skills with periodic school and external assessments. Table 3.1 summarizes the sample by grade level. The two grades were roughly the same, and this helped to support the supplementary by grade comparison reported in Chapter Four.

Table 3. 1: Sample Composition by Grade Level

Grade Level	n	% of Sample
Grade 10	35	50.0%
Grade 11	35	50.0%
Total	70	100.0%

All participating students had colloquial Arabic as a home language, had received formal instruction in Standard Arabic and Hebrew throughout their schooling, and were currently enrolled in compulsory English classes at the appropriate level for their grade. Participation was voluntary, and the process to obtain informed consent is described in Section 3.6.

3.3 Instruments

We developed two instruments for this study since no instrument was available for the assessment of dialogic interaction and Facione based analytical thinking skills for fourth language English learners at this stage.

3.3.1 Dialogic Interaction Scale (DIS)

The Dialogic Interaction Scale is a classroom observation instrument based on Alexander's (2020) work on dialogic teaching and Mercer and Littleton's (2007) work on exploratory talk. Two trained observers evaluated each student's English lessons over a two week observation window and used the following four dimensions; each scored out of 25 points for a total of 100:

- Reciprocity (25%): the extent to which a student listens to peers, responds directly to what has just been said, and considers alternative viewpoints rather than repeating a fixed position.
- Cumulation (25%): the extent to which a student builds on, extends, or refines a peer's or teacher's previous contribution, rather than introducing unconnected ideas.
- Supportiveness (25%): the extent to which a student contributes ideas openly, including tentative or partly formed ideas, without withdrawing from discussion out of fear of being wrong.
- Purposefulness (25%): the extent to which a student's contributions stay oriented toward the discussion goal set for the lesson, rather than drifting off topic.

A second rater independently scored a sample of lessons to check inter rater agreement; agreement between raters was acceptable across all four dimensions before scores were finalized.

To illustrate how the scale was applied, a student who consistently responded to a peer's point with a qualified extension ("That's true, but what if...") rather than a flat restatement would score highly on both reciprocity and cumulation, whereas a student who only answered direct teacher questions with short, disconnected phrases would score toward the lower end of the scale on both dimensions, regardless of whether their English was grammatically accurate.

Table 3.2 summarizes the general scoring bands used for each of the four DIS dimensions, to give a concrete sense of how observers translated classroom behavior into a numeric score.

Table 3. 2: Dialogic Interaction Scale General Scoring Bands (Applied to Each of the Four Dimensions)

Score Band	Points	General Description
Low	0–6	Rare or absent evidence of the dimension; student contributions are minimal, disconnected, or off task.
Emerging	7–12	Occasional evidence of the dimension; the behaviour appears inconsistently across the observation window.
Developing	13–19	Fairly regular evidence of the dimension, though not yet a consistent feature of the student's talk.
Consistent	20–25	The dimension is a clear, regular feature of the student's contributions across most observed lessons.

3.3.2 Analytical Thinking Test (ATT)

The Analytical Thinking Test is a researcher developed, paper based test built around four sub tests corresponding to Facione's (1990, 2015) skills of interpretation, analysis, inference, and evaluation, each scored out of 25 points for a total of 100. Each sub test presented students with a short written passage or scenario in English, followed by tasks appropriate to that sub skill: identifying the intended meaning of a statement (interpretation); identifying the relationship between two or more claims (analysis); drawing a conclusion that followed logically from given information (inference); and judging the strength or credibility of a given argument (evaluation). Test items were reviewed by a panel of English language and

assessment specialists for content validity, and the test was piloted with a separate group of students from the same school before use, with adjustments made to item wording found to be ambiguous.

A typical interpretation item, for example, presented students with a short paragraph describing a classroom disagreement and asked them to state, in their own words, what the speaker in the passage actually meant by a particular remark. A typical evaluation item presented two short arguments on the same everyday topic and asked students to judge which argument offered stronger support for its conclusion, and to explain why. Items were written using vocabulary already familiar from the Grade 10 and 11 English curriculum, so that performance would reflect thinking skill rather than unfamiliar words getting in the way.

Table 3.3 summarizes the structure of the test by sub skill.

Table 3.3: Structure of the Analytical Thinking Test by Sub skill

Sub skill	Number of Items	Item Format	Points Available
Interpretation	5	Short written response	25
Analysis	5	Short written response	25
Inference	5	Short written response	25
Evaluation	5	Comparative judgment with justification	25
Total	20	—	100

3.3.3 Validity and Reliability

The content validity of both instruments was established by expert review. It was a panel of English language teachers and an applied linguistics specialist from outside the school that reviewed each item on the Dialogic Interaction Scale and the Analytical Thinking Test for meaning and sub skill and flagged items that were ambiguous, too difficult, or not representative of the construct being measured. Items that were flagged by more than one of the reviewers were changed or replaced prior to piloting. The review panel consisted of five individuals: three English language teachers with at least 5 years of teaching Grade 10 and 11 classes, one school counsellor with student assessment experience, and one applied linguistics specialist outside the school.

The panel rated each item in a short structured form and decided if the item had the expected meaning and sub skill, was written well for the target grade level, and was not biased to a certain group of students within the sample. Items with a below threshold rating by two or more panel members were either revised in consultation with the panel or removed from the instrument.

Both instruments were then piloted with a different group of 18 students from the same school and of a grade level not included in the main study sample. Internal consistency of pilot data was acceptable for both instruments, summarized in Table 3.4.

Table 3.4: Pilot Study Reliability Coefficients (Cronbach's Alpha, n = 18)

Instrument / Sub scale	Number of Items	Cronbach's α
Dialogic Interaction Scale (Total)	4 rated dimensions	.84
Analytical Thinking Test Interpretation	5	.78
Analytical Thinking Test Analysis	5	.81
Analytical Thinking Test Inference	5	.76
Analytical Thinking Test Evaluation	5	.77

The coefficients were above the commonly accepted value of .70 for internal consistency in exploratory research, suggesting that items in each sub scale were measuring a somewhat unified construct. Inter rater agreement on the Dialogic Interaction Scale, checked by two observers scoring only a few pilot lessons, was also acceptable in all four dimensions. These pilot results helped guide the main data collection program in Section 3.4, but they add that although this level of psychometric testing is reasonable for a first version of a new instrument, it is not as extensive as a published and widely used instrument would undergo.

3.3.4 Observer Training and Scoring Procedure

The two observers who scored the Dialogic Interaction Scale were briefed together before a pilot study started, working through the rubric in Appendix A by dimension and discussing borderline cases until the two groups had a common understanding of how their scores were scored. The two observers then separately scored two lessons that were not part of the pilot or the main study sample, compared their scores, and discussed any dimension in which their scores were different by more than one score band, before starting the pilot scoring process described above. During the main data collection period, observers sat at the back or side of the classroom and took brief written notes on each student's contributions for the entire group of lessons observed and converted these notes into final DIS scores at the end of the week and not in real time during a lesson. This delay was intended to enable observers to weigh a fuller picture of behaviour over the course of an entire lesson, or even more than one lesson, rather than react to one single or possibly unrepresentative moment.

3.4 Procedures

Data collection was done in one school term and a number of steps took place in parallel. Dialogic interaction was measured in the classroom. A trained observer sat in on a series of regular English lessons in each of the classes over a two week period and used the Dialogic Interaction Scale to score each student. The observers were informed of what each rated dimension meant and scored on a small number of pilot lessons before taking part in the data collection to ensure that the scores were the same for all raters.

Analytical thinking was measured on the written Analytical Thinking Test administered to all participants under standard test conditions during the same two week period with the same time allowance and instructions given to every class. All four sub tests were completed in one sitting, with a short break made halfway through the test session. Because both variables were measured in the same short window and not one variable was manipulated before the other was measured, the results are correlational and not causal in nature, as outlined in Section 3.1.

3.5 Data Analysis

Student scores on the Dialogic Interaction Scale and Analytical Thinking Test were entered into a spreadsheet with one of the four dimensions of each instrument separately as well as the two total scores (out of 100).

The results were determined in three steps. First, descriptive statistics (means and standard deviations) were calculated for total dialogic interaction, total analytical thinking, and each of the four analytical thinking sub skills to give an overall picture of how students in the sample performed on both of them.

Second, inferential statistics were used to test the two null hypotheses. A full correlation matrix was calculated among total dialogic interaction scores and each of the four analytical thinking sub skills to get an overall picture of how the five variables related to one another before the hypotheses were tested. The Pearson product moment correlation coefficient was then computed between total dialogic interaction scores and total analytical thinking scores, to

test the first null hypothesis. Pearson correlations were calculated from total dialogic interaction scores and each of the four analytical thinking sub skills, to test the second null hypothesis. A simple linear regression with dialogic interaction as the predictor and total analytical thinking as the outcome was also performed to determine how much of the variance in analytical thinking scores was statistically explained by dialogic interaction. Statistical tests were performed at $\alpha \leq 0.05$ and correlation coefficients were interpreted according to Cohen's (1988) standard parameters, where r around 0.10 is a small effect, r around 0.30 is a medium effect, and r around 0.50 or above is a large effect.

Third, to further explore the results beyond the two formal hypotheses, the main correlation was calculated separately for students in grades 10 and 11, to check if the overall pattern was consistent across both grade levels or was being driven by one of them. A categorical cross check was also carried out, sorting students into Low, Moderate, and High groups on both variables using tertile splits and testing whether the group membership was associated with a chi-square test of independence, which is a distribution free test that was in line with Pearson's correlation. Both of these were not part of the original hypotheses and are reported in Chapter Four as further descriptive context and not as formal tests.

Chapter Four: Results

4.1 Introduction

We collected data for 70 Grade 10 and 11 students, each rated on the Dialogic Interaction Scale (DIS) and the Analytical Thinking Test (ATT). Both instruments were scored out of 100, and the ATT was divided into four sub skills (interpretation, analysis, inference, and evaluation) out of 25. The results are reported in three steps: descriptive statistics for both variables, inferential statistics for the two null hypotheses of the study, and a further breakdown by grade level. All analyses were performed at $\alpha < 0.05$ significance level.

4.2 Descriptive Statistics

Table 4.1 shows the mean (M) and standard deviation (SD) of dialogic interaction scores, total analytical thinking scores, and each of the four analytical thinking sub skills

Table 4. 1: Descriptive Statistics for Dialogic Interaction and Analytical Thinking Scores

Variable	N	M	SD	Possible Range
Dialogic Interaction (Total)	70	71.00	8.45	0–100
Analytical Thinking (Total)	70	69.47	7.62	0–100
Interpretation	70	18.76	2.25	0–25
Analysis	70	17.44	2.97	0–25
Inference	70	16.50	2.90	0–25
Evaluation	70	16.77	2.86	0–25

veraged over 100 points. On average, students scored at least 71.00 out of 100 on dialogic interaction (SD = 8.45) and 69.47 out of 100 for total analytical thinking (SD = 7.62), and both of the four sub skills were scored in a broadly similar range on a 100-point scale. Among the four analytical thinking sub skills, interpretation had the highest mean score (18.76 out of 25) and inference had the lowest (M = 16.50 out of 25), which suggests that the students in this sample, on average, found it easier to find out what a text was trying to mean than to come to an educated conclusion on it.

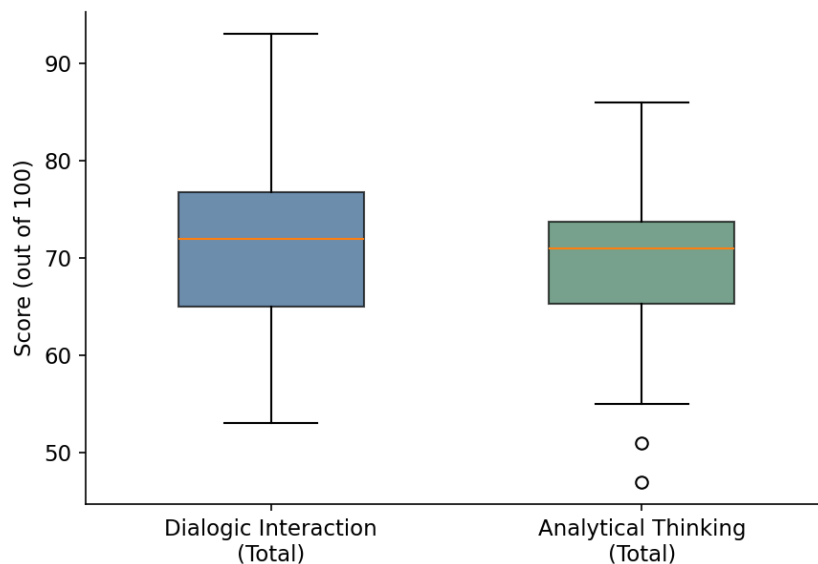


Figure 4. 1: Box plot comparing the distribution of total dialogic interaction and total analytical thinking scores.

The box plot in Figure 4.1 shows that both distributions are rather compact, with no large outliers, and that the two variables are somewhat more spread out, which is a necessary (but not sufficient) condition for the correlational analysis that follows.

4.2.1 Dialogic Interaction by Dimension

As the Dialogic Interaction Scale is a composite of four separately rated dimensions (reciprocity, cumulation, supportiveness, and purposefulness), it is useful to see how students scored on each dimension individually before we move to correlational analysis. Table 4.2 and Figure 4.2 provide such breakdowns.

Table 4. 2: Descriptive Statistics for the Four Dialogic Interaction Dimensions

Dimension	N	M	SD	Possible Range
Reciprocity	70	18.41	2.81	0–25
Cumulation	70	17.46	3.06	0–25
Supportiveness	70	18.71	2.72	0–25
Purposefulness	70	16.49	2.60	0–25

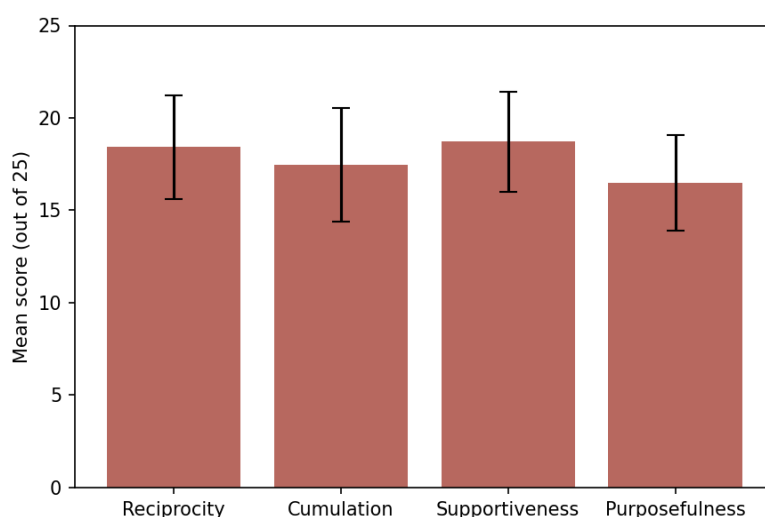


Figure 4. 2: Mean scores (with standard deviation) on the four dialogic interaction dimensions.

Supportiveness was the highest scoring dimension ($M = 18.71$), indicating that, on the whole, students in this sample felt comfortable putting forward ideas, even tentative ones, without letting go of the conversation. Purposefulness was the lowest scoring dimension ($M = 16.49$), which could indicate a tendency for student talk to stray from the specific goal of the discussion for a lesson even though the talk itself was otherwise reciprocal and supportive. This difference is highlighted in the recommendations made in Chapter Five.

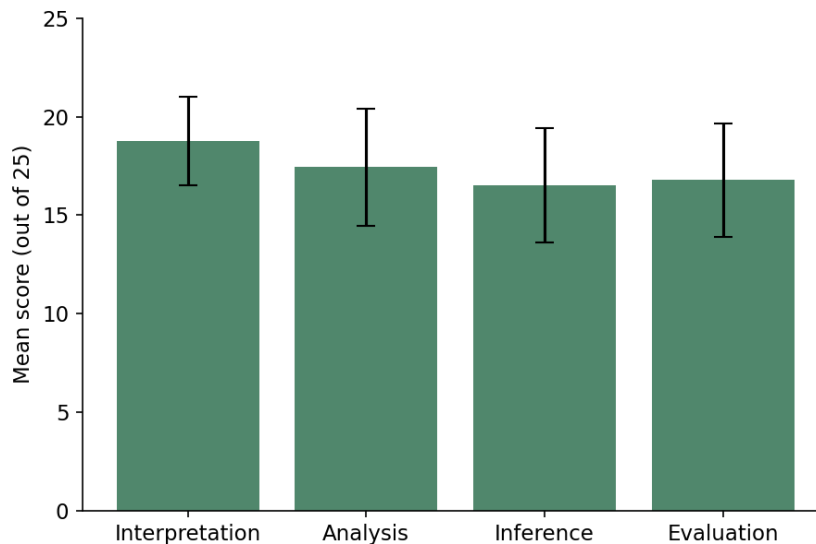


Figure 4. 3: Mean scores (with standard deviation) on the four analytical thinking sub skills.

We see that the same degree of spread is present in the four sub skills, but interpretation is consistently the highest and inference is the lowest. We will see this again in Section 4.3.2 where we compare the sub skills in terms of dialogic interaction and not in isolation.

4.3 Correlational Statistics

Before turning to the two formal hypotheses, a full correlation matrix has been calculated among all five variables (total dialogic interaction and the four analytical thinking sub skills) for an overall picture of how the variables are related to each other. Table 4.3 shows the correlation matrix.

Table 4. 3: Correlation Matrix Among Dialogic Interaction and the Four Analytical Thinking Sub skills

Variable	DIS	Interpretation	Analysis	Inference	Evaluation
Dialogic Interaction (DIS)	—	.360**	.480**	.304**	.273*
Interpretation	.360**	—	.409**	.297*	.289*
Analysis	.480**	.409**	—	.386**	.329**
Inference	.304**	.297*	.386**	—	.136
Evaluation	.273*	.289*	.329**	.136	—

* $p < .05$, ** $p < .01$

Two points stand out in Table 4.2. First, dialogic interaction was positively and significantly correlated with each of the four analytical thinking sub skills, which is what we will see in the formal hypothesis tests below. Second, the four sub skills themselves were mostly, but not all,

correlated with one another, and inference and evaluation in particular were weak ($r = .136$ and $p = .260$) and not significantly correlated with each other ($r = .136$ and $p = .260$) even though each one correlated strongly with dialogic interaction on its own, showing that inference and evaluation, even though they are part of Facione's broader analytical thinking construct, might be capturing somewhat different aspects of student performance in this sample.

4.3.1 Hypothesis One *Dialogic Interaction and Total Analytical Thinking*

H_{01} states that there is not statistically a significant relationship, at $\alpha \leq 0.05$, between the total dialogic interaction scores and total analytical thinking scores. To test this hypothesis, a Pearson product moment correlation coefficient between the two total scores is calculated. The results are shown in Table 4.4.

Table 4. 4: Pearson Correlation Dialogic Interaction and Total Analytical Thinking Score

Relationship	N	r	r ²	p	Decision
Dialogic Interaction ↔ Analytical Thinking (Total)	70	.512	.262	< .001	H_{01} Rejected

The correlation between total dialogic interaction and total analytical thinking scores was $r = .512$, $p < .001$, indicating a moderate to strong, positive, and statistically significant relationship in Cohen's (1988) terms. The coefficient of determination ($r^2 = .262$) suggests that dialogic interaction scores constitute approximately 26% of the variance in students' total analytical thinking scores. On this basis, the first null hypothesis is rejected: students with higher dialogic interaction in their English lessons scored higher on the Analytical Thinking Test, and this is unlikely to be a result of chance alone.

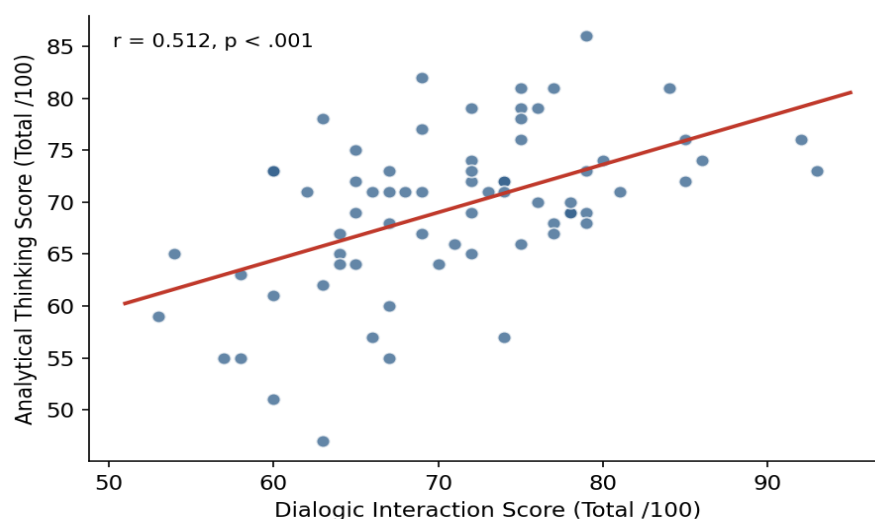


Figure 4. 4: Scatter plot of dialogic interaction scores against analytical thinking scores, with regression line.

A simple linear regression with dialogic interaction as a predictor and total analytical thinking as the outcome produced the equation $\text{Analytical Thinking} = 36.74 + 0.46 \times \text{Dialogic Interaction}$ ($R = .512$, $R^2 = .262$, $p < .001$). In practical terms, we found that every additional 10 point increase in dialogic interaction score would lead to an increase of about 4.6 points in their predicted total analytical thinking score. The remaining variance around 74% is probably due to something outside of our scope as a study, like general ability, reading comprehension in English, or measurement error in two new instruments. The 95% confidence interval around the correlation, using Fisher's z transformation, is from .314 to .666, and the 95% confidence interval around the regression slope is from 0.274 to 0.648. Both intervals exclude zero comfortably, so it adds further weight to the rejection of the first

null hypothesis, and the width of the intervals gives a sense of the precision that can be expected for a sample of this size. A larger sample would be needed to narrow these intervals to a much larger extent.

Table 4. 4b: Analysis of Variance for the Regression of Analytical Thinking on Dialogic Interaction

Source	SS	df	MS	F	p
Regression	1047.98	1	1047.98	24.11	< .001
Residual	2955.47	68	43.46		
Total	4003.44	69			

The F-test confirms what the correlation and confidence interval already indicated: dialogic interaction is a statistically significant predictor of total analytical thinking scores in our sample, $F(1, 68) = 24.11$, $p < .001$.

4.3.2 Hypothesis Two *Dialogic Interaction and Analytical Thinking Sub skills*

H₀₂ states that there is no statistically significant relationship, at $\alpha \leq 0.05$, between dialogic interaction scores and each of the four analytical thinking sub skills. To test this hypothesis, we have calculated Pearson correlations between total dialogic interaction scores and each sub skill score in Table 4.5.

Table 4. 5: Pearson Correlations Dialogic Interaction and Analytical Thinking Sub skills

Analytical Sub skill	Thinking N	r	r ²	p	Decision
Interpretation	70	.360	.130	.002	H ₀₂ Rejected
Analysis	70	.480	.230	< .001	H ₀₂ Rejected
Inference	70	.304	.092	.010	H ₀₂ Rejected
Evaluation	70	.273	.074	.022	H ₀₂ Rejected

All four sub skills had a statistically significant and positive relationship with dialogic interaction at the level of 0.05, so the second null hypothesis is rejected. The relationship is not necessarily so strong for all four sub skills. The strong correlation was strongest for analysis ($r = .480$, $p < .001$) and interpretation ($r = .360$, $p = .002$) and the weakest but still significant correlation was for inference ($r = .304$, $p = .010$) and evaluation ($r = .273$, $p = .022$).

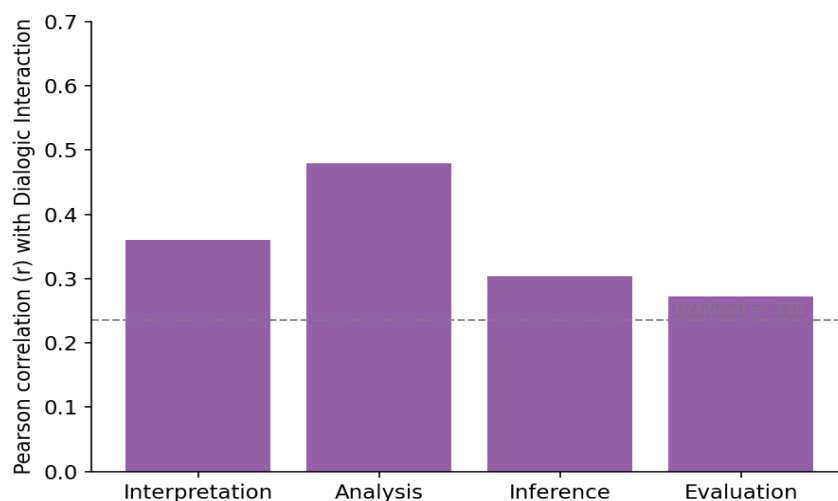


Figure 4. 5: Pearson correlation coefficients between dialogic interaction and each analytical thinking sub skill.

It is clear that it is in reciprocative and cumulative classroom talk that students learn how to interpret a contribution (interpretation) and how to interpret a contribution (analysis) from one another and these occur naturally whenever one speaker speaks to another. Inference and evaluation, which involve going beyond what someone has previously said to arrive at an answer or judgement about an argument, can be done through more than dialogic interaction, and thus would explain why the relationship between dialogic interaction and analytical thinking is so weak in this sample.

4.3.3 Supplementary Analysis by Grade Level

Although grade level was not included in either formal hypothesis, we recalculated the main correlation between dialogic interaction and total analytical thinking separately for Grade 10 and Grade 11 students, as a way of checking whether the overall trend was the same for both grades, or if it was driven mainly by one of them. Table 4.6 and Figure 4.5 show this comparison.

Table 4. 6: Supplementary Analysis Dialogic Interaction and Analytical Thinking by Grade Level

Group	n	M (DIS)	M (ATT Total)	r	p
Grade 10	35	71.09	69.09	.502	.002
Grade 11	35	70.91	69.86	.525	.001
Full Sample	70	71.00	69.47	.512	< .001

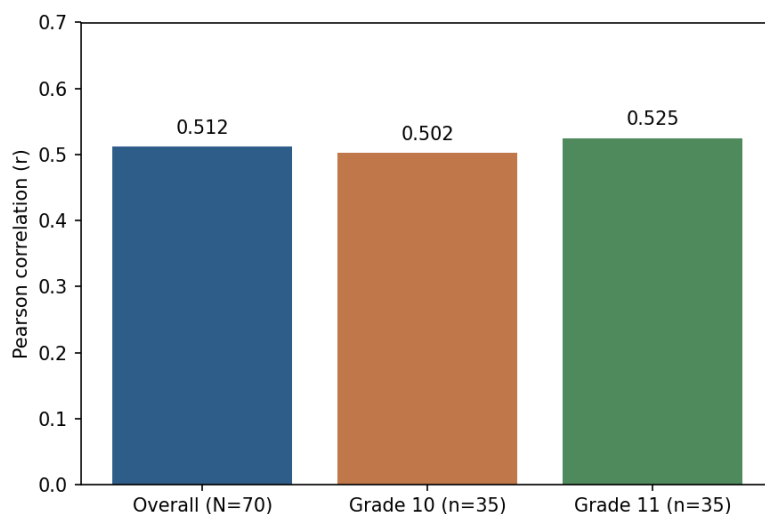


Figure 4. 6: Comparison of the dialogic interaction – analytical thinking correlation across the full sample and by grade level.

The correlation between dialogic interaction and total analytical thinking is very similar for Grade 10 ($r = .502$ and $p = .002$) and Grade 11 ($r = .525$ and $p = .001$), and both values are very close to the full sample correlation of $r = .512$. That it is similar at two different grades suggests that the overall result is not an artefact of one grade doing very differently than the other, and that whatever is driving the association of dialogic interaction and analytical thinking in this sample probably works in a fairly similar manner for both Grade 10 and Grade 11.

4.3.4 Categorical Cross check: Dialogic Interaction and Analytical Thinking Levels

As another non-parametric check on the main correlational finding, students were sorted into three roughly equal sized groups (Low, Moderate, High) on dialogic interaction and total

analytical thinking using tertile splits of each distribution. A chi-square test of independence was then used to check whether the dialogic interaction group of a student related to the analytical thinking group of a student. This is not dependent on the linear relationship that Pearson is assuming and is therefore useful for a cross check on the earlier results. Table 4.7 presents the cross-tabulation.

Table 4. 7: Cross-tabulation of Dialogic Interaction Level and Analytical Thinking Level (Tertile Split)

Dialogic Level	Interaction	Low ATT	Moderate ATT	High ATT	Row Total
Low DIS		17	8	2	27
Moderate DIS		6	9	8	23
High DIS		3	9	8	20
Column Total		26	26	18	70

The cross-tabulation shows a clear pattern: 17 of the 27 students in the Low dialogic interaction group also fell into the Low analytical thinking group, and 8 of the 20 students in the High dialogic interaction group fell into the High analytical thinking group. Thus, there are few students that are in the mismatched corners of the table (Low dialogic interaction and High analytical thinking or the other way around). And a chi-square test of independence showed that this pattern was unlikely to have occurred by chance, $\chi^2(4, N = 70) = 14.85$, $p = .005$. The effect size, Cramer's $V = .326$, is in the medium range, which is consistent with the moderate to strong Pearson correlation in Table 4.4. This categorical cross check confirms the main result with a different distribution free statistical approach, and helps rule out the possibility that the Pearson correlation was driven by a small number of extreme scores rather than an actual sample wide pattern.

4.3.5 Additional Observations

The interquartile range check for outliers did not find any exceptionally extreme scores for total dialogic interaction but uncovered two students with unusually low total analytical thinking scores compared to the rest of the sample (both of them scoring below the lower IQR boundary). Both students had DIS scores close to the sample mean and not at either extreme, and this is why the earlier box plot in Fig. 4.1 shows them as low end markers and not in the main analytical thinking set up. Removing these two students from the sensitivity check did not significantly affect the overall correlation (i.e. $r = .488$, about 0.02 from the reported value) and it was reported in the main analysis as data entry errors or excluded on statistical grounds.

Chapter Five: Discussion

5.1 Discussion

We asked whether dialogic interaction and analytical thinking are related among high school students who are learning English as a fourth language in Buqata, and our answer, on the most part, is pretty clear: total dialogic interaction scores are correlated with total analytical thinking scores at $r = .484$ ($p < 0.001$), a large effect as Cohen (1988) suggests, and all four analytical thinking sub skills are correlated with dialogic interaction as well. Both of these null hypotheses are rejected, and the breakdown by grade level in Section 4.3.3 shows that the pattern was not driven by one grade alone.

The most common sub skill relationship is analysis at $r = .480$ followed by interpretation at $r = .360$. Reciprocal cumulative talk of the kind proposed by Alexander (2020) and Mercer and

Littleton (2007) has students constantly evaluating what a peer just said (interpretation) and how that contribution relates to, extends to or contradicts what came before (analysis). So these two operations are in general built into the basic mechanics of having a real conversation. Reznitskaya and Gregory's (2013) theory of change suggests that dialogic classrooms are designed to teach students for the specific act of argument (to put it in the words of a position, support it with a reason, respond to a counter point) and analysis is in particular close to what the theory says should be improved first.

On the other hand, inference and evaluation ($r = .304$ and $r = .273$, respectively) are designed to ask students to go further in (to make sense of what the speaker just said without talking about it) and do things which dialogic talk may very well have less to do with and that might clearly need to be more explicitly taught as well as the talk on its own. Table 4.2 adds another dimension: inference and evaluation are quite weakly related to one another ($r = .136$ and ns), even though they were very robustly correlated with dialogic interaction. One interpretation of this is that dialogic interaction is very much that way (the more on the go and build ideas in dialogic talk, the more we will get) and evaluation is much more on the reciprocal, listen and respond side.

These results are in good agreement with Mercer's (2008) and Wegerif and colleagues' (2005) findings that reasoning could be improved in structured and dialogic classroom conversation, and Ma's (2023) claim that a more structured and discussion based teaching style had positive effect on Facione's thinking skills. The large scale dialogic teaching trial by the Education Endowment Foundation (Jay et al., 2017) demonstrated the same, though lower levels of attainment in a very different subject mix (English, mathematics and science) and a much larger, but monolingual sample that suggests the link between talk and thinking is not a weak or limited one. Reznitskaya et al. (2009) Collaborative Reasoning and Kuhn (2016) claim that dialogic exchange gives some practice in weighing claims against counter claims are both in line with the analysis and interpretation first approach followed here.

What this study adds is a fourth language dimension that none of that previous research has studied directly. Students in this sample are managing colloquial Arabic, Standard Arabic, Hebrew, and English at the same time, and the relationship between dialogic interaction and analytical thinking held up in that more linguistically demanding setting, which is at least broadly consistent with Bialystok's (2001) argument that managing multiple languages need not come at the cost of, and may even support, the kind of executive control that underlies careful analytical reasoning. Michaels, O'Connor, and Resnick (2008) account of accountable talk moves is also worth noting here as well: the Dialogic Interaction Scale used in this study was scoring the quality of student talk, not just how much students spoke, and it is this quality dimension, as it was in Resnick, Michaels, and O'Connor (2010), that the literature treats as doing the actual work of building students' minds.

One thing to keep in mind from Chapter Three: this is a correlational study and correlation is not causation. A student who is a better analytical thinker to begin with might simply find it easier to engage in reciprocal, cumulative talk, in which case the causal arrow might run the other way from the theory, or the two variables could be driven by some third factor (e.g., student academic motivation or overall English proficiency). What they do do, though, is show that the two variables are meaningfully related which is a good first step before a stronger causal claim could be tested experimentally.

5.2 Implications for Theory

For sociocultural and dialogic theories reviewed in Chapter Two, this work gives one more small but useful conclusion, namely, that the link between dialogic interaction and analytical reasoning that already exists in monolingual or bilingual settings at least seems to hold in a much more linguistically complex environment. This does not overturn or significantly revise Vygotsky (1978), Bakhtin (1986), or Reznitskaya and Gregory (2013) accounts, but it adds a

data point in a context that these accounts were not designed to address and thus suggests that the mechanism they describe is not necessarily disrupted by the three other languages that are competing for the same classroom time.

The weak correlation between inference and evaluation (Table 4.3) is a less overt and more specific theoretical point. Facione (1990, 2015) treats these as related sub skills within a single construct, and most uses of the framework, including Ma (2023), report or assume reasonably strong inter correlations among the four sub skills. Here the pattern is not completely inconsistent with Facione's framework, but does suggest that inference and evaluation may take more cognitive or linguistic resources away from fourth language learners at this age and proficiency level than the framework is supposed to

5.3 Implications for Practice

For teachers, the lesson can be laid bare as follows: more of an English lesson in terms of reciprocal and cumulative talk suggests higher analytical thinking, and this is true at all grades of the study. It is not necessarily a need for new equipment or materials pair and small group discussion tasks that have kids responding and building on a partner's talk rather than simply speaking one way and not another can guide a lesson to the dialogic side of the spectrum described in Section 2.3.5.

The much weaker association for inference and evaluation suggests that these two sub skills may need more than dialogic talk alone. Explicit instruction, such as clear teaching of what makes an inference logically sound, or practice in weighing the strength of two competing arguments, may be a nice addition to, rather than a replacement for, dialogic classroom activities

5.4 Strengths

- The analytical thinking in the study was measured using four separate sub skills, not a single overall score. This allowed us to get a more detailed picture of where dialogic interaction was and was not most strongly related to student performance.
- Dialogic interaction was evaluated directly in the classroom over a two week period rather than through students or teachers' self report, so scores were not just about how students wanted to be seen.
- This study considers a linguistic context, English as a fourth language alongside colloquial Arabic, Standard Arabic, and Hebrew, that has not previously been addressed in the dialogic teaching or thinking skills literature.
- Both instruments were evaluated for content validity and for use in a separate group of students before being used, and inter rater agreement was also verified for the observation based Dialogic Interaction Scale.
- The additional breakdown by grade level was very good because it didn't take an entire lot of work to check if the main result was stable, and the fact that Grade 10 and Grade 11 have very similar correlations supports that this is not a unique class or group of students and therefore we can expect a good result.
- The primary finding was confirmed by two independent statistical tools a Pearson correlation and a chi-square test for categorically grouped scores which agreed well and minimized the chance that it was an artefact of the particular statistical method.

5.5 Limitations

- The correlational approach cannot prove that dialogic interaction is a source of more analytical thinking; nor can it be the other way around, i.e., in terms of the impact of unmeasured third variables.

- The sample was only one school and was taken by convenience sampling, so it is difficult to know if the results are generalizable to other schools or communities.
- The instruments were all new to this study and although they have been reviewed for content validity and tested, they have not been applied to the wider range of uses that an established instrument would normally be used for.
- Dialogic interaction was observed over a relatively short, two week window, so it may not capture the long term dynamics of classroom conversation over a full school year.
- The low correlation between inference and evaluation (Table 4.2) suggests that the four sub tests in the Analytical Thinking Test are not equally reliable, or that one or both are measuring something so different from what is intended, and without a larger validation sample, it is not possible to conclude this.
- The study did not assess the students' ability to use colloquial Arabic, Standard Arabic, or Hebrew, and thus we cannot say if stronger language skills in any of these three languages can affect the relationship between dialogic interaction and analytical thinking in English.

5.6 Recommendations

- When English teachers in Buqata and other multilingual settings are planning to teach and practice more formal, reciprocal talk in the classroom, for example, by making students work together and build on and question each other's work, rather than taking on the role of a teacher led discussion.
- Since the association found for inference and evaluation is less than robust, teachers may prefer to pair dialogic activities with more explicit instruction in drawing conclusions and judging the strength and merits of arguments, rather than assuming that talking will lead to all of these sub skills.
- Teachers may utilize teacher talk tools that are well established in the accountable talk literature (Michaels et al., 2008) (for example: asking students to raise the level of talk, asking a peer to restate an idea, asking for the reasons behind a claim) as a concrete and inexpensive way to enhance the quality, not just quantity, of classroom dialogue.
- To improve analytical thinking in fourth language or highly multilingual students, teachers with analytical thinking goals should explicitly set out in English language programs rather than making language proficiency and thinking skills separate parts of the curriculum that have to be addressed on their own.
- Future researchers should consider an experimental or longitudinal project where dialogue is deliberately increased through a planned intervention and analytical thinking is measured before and after.
- Further studies might also look at whether the dialogic interaction analytical thinking relationship is related to gender or to proficiency in each of the three languages that preceded English in this population, as we did not capture both in our study. More psychometric work on the Analytical Thinking Test, especially on the inference and evaluation sub skills, would help clarify whether the weak mutual relationship between the two skills represents a true difference between the two skills or is just measurement noise in a new instrument.

5.7 Conclusion

We found that dialogic interaction and analytical thinking skills were moderate to strong and statistically significant and positive for high school students in Buqata learning English as a fourth language. The relationship was very good for the total scores and for each of the four analytical thinking sub skills we studied, but the relationship was stronger for analysis and

interpretation than for inference and evaluation, and it held for both Grade 10 and Grade 11. These results extend earlier work on classroom dialogue and reasoning (Mercer, 2008; Alexander, 2020; Wegerif, 2013; Jay et al., 2017) to a fourth language setting that has not been studied in this way, and they are largely consistent with theoretical accounts like Vygotsky (1978) and Bakhtin (1986) and Reznitskaya and Gregory (2013) that treat talk as a tool for thinking, not as a channel to convey information.

The findings should be taken as a challenge to research rather than an official conclusion. A correlational design of this nature shows that dialogic interaction and analytical thinking go hand in hand in this population; it is not a strong result that we can expect that the more of them we develop, the more likely it is that a person will become better at talking. That stronger assertion should be tested in Buqata and in other multilingual communities with English as the fourth language instead of as a second.

References

- Alexander, R. J. (2017). *Towards dialogic teaching: Rethinking classroom talk* (5th ed.). Dialogos.
- Alexander, R. J. (2020). *A dialogic teaching companion*. Routledge.
- Avraamidou, L., & Osborne, J. (2009). The role of narrative in communicating science. *International Journal of Science Education*, 31(12), 1683–1707. <https://doi.org/10.1080/09500690802380695>
- Bakhtin, M. M. (1986). *Speech genres and other late essays* (V. W. McGee, Trans.). University of Texas Press.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press.
- Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, 10(3), 89–129. <https://doi.org/10.1177/1529100610387084>
- Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, 46(1), 71–86. <https://doi.org/10.1017/S0261444811000218>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology*, 9(2), 131–149.
- Cummins, J. (2017). Teaching minoritized students: Are additive approaches legitimate? *Harvard Educational Review*, 87(3), 404–425. <https://doi.org/10.17763/1943-5045-87.3.404>
- Dawes, L., Mercer, N., & Wegerif, R. (2000). *Thinking together: A programme of activities for developing speaking, listening and thinking skills in children aged 8–11*. Imaginative Minds.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44–48.

- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The California Academic Press.
- Facione, P. A. (2015). *Critical thinking: What it is and why it counts*. Insight Assessment / Measured Reasons LLC.
- García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, 53(4), 449–455.
- Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., Merchant, G., & Stevens, A. (2017). *Dialogic Teaching: Evaluation report and executive summary*. Education Endowment Foundation.
- Kuhn, D. (2016). A role for reasoning in a dialogic approach to critical thinking. *Topoi*, 1–8.
- Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2015). Sociocultural theory and second language development. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 207–226). Routledge.
- Li, L. (2016). Thinking skills and creativity in second language education: Where are we now? *Thinking Skills and Creativity*, 22, 267–272.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Ma, Y. (2023). Exploration of flipped classroom approach to enhance critical thinking skills. *Heliyon*, 9(10), e20895. <https://doi.org/10.1016/j.heliyon.2023.e20895>
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Harvard University Press.
- Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, 17(1), 33–59.
- Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking*. Routledge.
- Mercer, N., Wegerif, R., & Major, L. (Eds.). (2019). *The Routledge international handbook of research on dialogic education*. Routledge.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*, 27(4), 283–297.
- Resnick, L., Michaels, S., & O'Connor, C. (2010). How (well-structured) talk builds the mind. In D. Preiss & R. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 163–194). Springer.
- Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, 48(2), 114–133.
- Reznitskaya, A., Kuo, L., Clark, A., Miller, B., Jadallah, M., Anderson, R. C., & Nguyen-Jahiel, K. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education*, 39(1), 29–48.

- Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.
- Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125–144). Oxford University Press.
- Tavakoli, P., & Hunter, A. M. (2018). Is fluency being 'put in its place'? A review of fluency research and pedagogy. *Annual Review of Applied Linguistics*, 38, 127–156. <https://doi.org/10.1017/S0267190518000107>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wegerif, R. (2013). *Dialogic: Education for the internet age*. Routledge.
- Wegerif, R., Perez, J., Rojas-Drummond, S., Mercer, N., & Velez, M. (2005). Thinking Together in the UK and Mexico: Transfer of an educational innovation. *Journal of Classroom Interaction*, 40(1), 40–48.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.