



Talking Science

Argument-Based Inquiry, Teachers' Talk Moves, and Students' Critical Thinking in the Classroom

Yilmaz Soysal¹ 

Accepted: 9 September 2020 / Published online: 30 September 2020

© Springer Nature B.V. 2020

Abstract

The paper reports an empirical study that examined relationships between the implementation quality of an argument-based inquiry (ABI) approach, teachers' talk moves, and students' critical thinking. Students were taught using an ABI approach in which they had opportunities to engage in critical thinking. However, this instructional approach was handled by a variety of teachers with differing capacities to implementing it. Systematic observation (talk move analysis) was therefore required to describe how quality of implementation affected students' critical thinking. Four middle school science teachers and 92 seventh-grade students participated. Videos of in-class implementations were examined through classroom discourse analysis. It was observed that the ABI approach seemed to stimulate student-led critical thinking when implemented with higher quality. Teacher-led talk moves were correlated with implementation quality and had the ability to evoke or hinder students' critical thinking. Instances of teachers evaluating students' responses based on scientific knowledge and transmission of knowledge tended to inhibit students' execution of higher-order critical thinking. Frequent use of less cognitively demanding talk moves, such as observe–compare–predict, did not seem to help students attain higher-order critical thinking. Four talk moves in particular (legitimizing, challenging, monitoring, evidencing) appeared to be associated with higher implementation quality, in turn promoting students' critical thinking. Communicating moves appeared to facilitate the initiation and maintenance of elaborate verbal exchanges. These encouraged students to perform higher-order critical thinking, though only in the presence of other talk moves (e.g., challenging, legitimizing, monitoring, and evidencing).

✉ Yilmaz Soysal
yilmazsoysal@aydin.edu.tr

¹ Faculty of Education, Department of Elementary Education, Istanbul Aydin University, Istanbul, Turkey

1 Introduction

Encouraging students to engage in critical thinking in the science classroom has been acknowledged as a major goal (Tiruneh et al. 2018). Promoting critical thinking is widely accepted as essential for the meaningful acquisition of science concepts and science process skills (e.g., Kuhn 1999; Tiruneh et al. 2017). Core components of critical thinking such as decision-making, inference, advanced clarification, analysis, evaluation, and explanation are intertwined with the thinking practices used in scientific inquiry. It is also recognized that elementary and middle school students may find both scientific and critical thinking to be especially challenging (Cavagnetto 2010; McDonald and McRobbie 2012). The empirical study reported in this paper explored how science teachers' implementation of argument-based inquiry (ABI) affected students' critical thinking. Students were taught using an ABI approach whereby they had opportunities to think critically while constructing arguments related to inquiry activities. The ABI approach was handled differently by individual teachers who displayed varying capacities to implementing it. The capabilities of the teachers were initially measured by a Reformed Teaching Observation Protocol (RTOP). In addition, systematic observation of teacher-led talk moves allowed fine-grained inferences to be made on the relationship between the effectiveness of teachers' implementation of ABI and the critical thinking operations of students.

2 Theoretical Framework

2.1 Critical and Scientific Thinking in Science Education

This paper assumes that interactions exist between scientific reasoning/thinking and critical thinking. Kind (2013) noted that scientific thinking has roots in history, philosophy, sociology, nature, and the practice of science itself. The roots of critical thinking, however, are in cognitive psychology, and cultivated by philosophy (Paul and Elder 2008). Scientific thinking is mostly explicated by referring to scientists' habits of mind, or technical ways of knowing (Kind 2013). Studies of the history of scientific thinking have proposed taxonomies of core practices in science inquiry: experimental exploration, probabilistic and statistical thinking, categorization and classification, mathematical deduction, and hypothetical modelling (Kind 2013). Faye (2014) and Kuhn (2002) also claim a relationship between scientific thinking and argument construction (coordination of theory (claim) and data to generate warranted reasoning to seek and interpret natural phenomena).

Erduran and Dagher (2014) distinguished two dimensions of nature of science: cognitive-epistemic and social-institutional. Given the interaction of scientific thinking and inquiry with critical thinking, this study focuses on the cognitive-epistemic dimension and the related processual skills of science (e.g., observation, comparing, projecting). The skills and practices involved in undertaking science professionally and in the classroom can be listed as taking alternative explanations into account, creativity in thought and experiment, intellectual rigor, peer evaluation, and social validation of ideas, resolving anomalies in data collected, dealing with counter arguments of others in the learning environment, critical examination of baseless claims, classification and/or category construction of ideas/events/objects for inductive reasoning, and theory-laden observation. (Faye 2014; Kind and Osborne 2016; Kind 2013).

Critical thinking is “reasonable reflective thinking focused on deciding what to believe or do” (Ennis 1993; p. 180). Glaser (1941) proposed three dimensions of critical thinking that are pertinent to the cognitive-epistemic dimension of scientific thinking. These are eagerness to consider claims in a thoughtful manner by excluding intuitive propositions, having methodological knowledge of thinking critically, and developing accountability to disciplinary norms of logical reasoning. Critical thinking by scientists entails data collection, analysis, and interpretation, to make decisions in cutting-edge situations (Paul and Elder 2008) or to deal with anomalous data (Faye 2014; Kind and Osborne 2016; Kind 2013). To undertake higher-order critical thinking, an individual may perform a variety of cognitive activities: analyzing, understanding, inferring, reasoning inductively and deductively, and legitimating the validity and reliability of data sources. All of these are associated with core practices of scientific thinking. This study seeks to examine the interaction between implementation quality of ABI and teacher-led talk moves, and how these pedagogical-discursive conditions stimulated student’s critical thinking about scientific questions.

2.2 The ABI Approach to Stimulating Critical Thinking

ABI is a student-centered approach which encourages learners to gather, analyze, and interpret data, and to express their thoughts on an in-class science inquiry, though not as the end-product of the activity (Cavagnetto and Hand 2012). In this approach, construction of a scientific argument is an embedded element. Students are expected to arrive at their own conclusions in an evidence-based manner. Critiquing is a core element, as students must present convincing arguments when defending their particular points of views (Cavagnetto and Hand 2012). Rigorous social negotiation of meaning is also a core epistemic practice, and students’ ideas are tested against an interpreted data set and logic (Norton-Meier et al. 2008). In a typical instance of ABI, language plays a critical role. Students are encouraged to live the language of science by acting like scientists, and the arguments they put forward are tools of inquiry whereby language is created and used for mediation (Cavagnetto 2010). In sum, during a typical ABI process, students must seek data and evidence to support their claims, substantiate their hypothetical propositions, protect their assertions from counter arguments, and either critique or confirm others’ arguments.

The ABI approach can be assumed to stimulate students’ critical thinking. Critical thinking is associated with argumentation, critical evaluation, and higher-order thinking, all of which are typically inherent to ABI. Some researchers (Andrews 2015; Soufi and See 2019) define critical thinking with reference to Toulmin’s (1958) argument model, in which individuals are expected to coordinate both data and claim to generate reasoned discourse. As Andrews (2015) notes, critical thinking and argumentation are closely related. As applied to argumentation, critical thinking can encompass comprehending assumptions, making assertions, presenting evidence, and drawing conclusions warranted by the evidence (Andrews 2015; Soufi and See 2019). The above processes are expected to be employed by students during an ABI exercise. Hypothetically, ABI can be treated as an instrumental tool to encourage students to think critically, but the approach must be implemented appropriately, with high quality.

2.3 Description of Implementation Quality in the Context of this Study

The hypothesis of the present study is that when an ABI approach is implemented authentically (with high quality), it may arouse students' critical thinking. The term authenticity is central. It has been widely acknowledged that authentic in-class talk-based teaching strategies are inherently, positively, and significantly instrumental in stimulating critical thinking. (McPeck 1990b; Moore 2011; Tiruneh et al. 2014). Such strategies include collaborative reasoning (e.g., Reznitskaya et al. 2009), collective argumentation (e.g., Conner et al. 2014a, b)), and dialogic teaching (e.g., Alexander 2008). A well-designed, purposefully implemented, and properly staged immersion intervention (e.g., ABI as a talk-based teaching strategy) serves to encourage students to engage in higher-order critical thinking (McPeck 1990a, b; Tiruneh et al. 2018; Tiruneh et al. 2017).

Authenticity implies a true or exact implementation of the ABI approach (with high quality), that prompts students to engage in higher-order critical thinking. In teaching science, implementation quality signifies the effectiveness with which a teacher implements in-class inquiry (Campbell and Erdogan 2005). Closely related to implementation quality is the process quality of classroom discourse, regulated mainly by teacher-led talks (Mameli and Molinari 2014; Molinari and Mameli 2013). It has been well argued that good in-class science inquiry produces positive results in terms of conceptual change and concept formation (Cairns and Aarepattamannil 2019; Geier et al. 2008). It may be inferred that better implementations of in-class science inquiry can generate positive results in terms of stimulating higher-order critical thinking. However, a given science teacher may lack the ability to implement an approach to in-class inquiry (e.g., ABI) which stimulates critical thinking.

Varying levels of quality in the implementation of ABI may have dramatic effects on critical thinking. Several scholars have attempted to test the effectiveness of immersive learning environments in stimulating critical thinking (small-group discussion, Stark 2012; teacher-led higher cognitively demanding questioning, Barnett and Francis 2012). However, the results are contradictory (significant effect, e.g., Bensley and Spero 2014; nonsignificant effect, e.g., Toy and Ok 2012). This inconsistency may be a result of authenticity, that is, whether a version of in-class instruction designed to stimulate critical thinking was implemented as intended and desired (Abrami et al. 2015; Tiruneh et al. 2014). Exploring implementation quality and its impact on stimulating students' critical thinking will thus help to address the above dilemma.

3 Justification for the Empirical Study

3.1 Two Measures of the Implementation Quality of ABI

3.1.1 The RTOP Scale and its Restricted Scope in Assessing Implementation Quality

To analyze the quality of implementation, Martin and Hand (2009) and Kilic (2016) used a specific version of the RTOP (Piburn et al. 2000), referring to thirteen particular descriptors (Martin and Hand 2009; pp. 24–25). In the RTOP, the primary criterion is the teacher's capacity to promote dialogic teaching by boosting student voice. Second is the teacher's role as a guide, model, and resource person, rather than a controller of knowledge or a primary knower/evaluator of ideas. Third is the teacher's pedagogic ability to drive scientific debate

among students by explicitly requiring evidence-based reasoning. In this study, the RTOP was used to measure implementation quality levels among participating teachers.

Martin and Hand (2009) and Kilic (2016) explored the quantification of implementation quality through the RTOP to provide preliminary or restricted insights into a teacher's adequacy as an implementer. In the present study, a fine-grained reexploration of implementation quality was therefore carried out by systematically observing teachers' talk moves. It seemed necessary to determine which underlying talk moves most effectively provoked higher-level critical thinking. It is argued that "researchers tend to attribute improvements in students' learning to the effectiveness of the sequence of teaching activities, giving little explicit attention to the teacher's role in staging those teaching activities" (Leach and Scott 2002, p. 115). In the literature, in-class teaching activities to promote critical thinking have mostly been implemented without reference to the talk which surrounds them (Leach and Scott 2002, 2003; Soysal and Yilmaz-Tuzun 2019). The position taken here is that the way in which the teacher works with students, and talks a scientific story into existence, is central to any teaching sequence (Leach and Scott 2002, p. 124). Although the typologies of activities used in science classrooms are of considerable importance, their quality can only be made apparent through explicit reference to teacher-led talk moves (Leach and Scott 2002, 2003; Soysal and Yilmaz-Tuzun 2019). Uniquely, this study recognizes the mediational function of teacher-led talks in terms of crystallizing the effects of particular teaching activities on student-led gains (e.g., critical thinking). It is strongly recommended that implementation quality should be characterized with concrete reference to the talk moves that enclose or materialize a given in-class teaching activity. Furthermore, students are not automatically able to perform (higher-order) critical thinking by themselves. Arousing their (critical) thinking requires specific types of teacher-led talk moves (e.g., Chin 2007). This study therefore aims to improve the scope of RTOP-based evaluations by undertaking talk move analysis to capture a more holistic picture of the relationship between implementation quality and critical thinking.

3.1.2 A more Analytical and Powerful Indicator of Implementation Quality: Teacher-Led Talk Moves

Talk move refers to discursive remarks by science teachers which create opportunities for students to make cognitive contributions (e.g., critical thinking operations) in classroom discourse (Tytler and Aranda 2015). For example, if a teacher asks a student to elaborate on his/her idea, an elicited answer may be provided (Chin 2007; Pimentel and McNeill 2013). This may provoke critical thinking (e.g., externalizing advanced clarifications). When a teacher rejects a student-led response and seeks prescriptive meaning, alternative points of views will not be interrogated (van Booven 2015). This may inhibit critical thinking operations such as hypothesizing or inductive and deductive reasoning (Chin 2007).

Talk moves may increase or decrease cognitive demands on students, thus modifying student-led cognitive contributions (Soysal 2019; Lefstein et al. 2015). With the question ("What is an atom?"), a science teacher may demand less cognitive effort from students (retrieving knowledge from long-term memory). By contrast, a science teacher who changes his/her talk moves ("What is your comment on your classmate's definition of atom, let's evaluate that?") may create higher cognitive demands, stimulating student-led critical thinking through, e.g., argument examination.

More dialogic or open-ended talk moves may foster student voices and increase students' speaking time. Their articulations may then be elaborated, an indicator of productive classroom talk (Martin and Hand 2009; van Booven 2015). The contingency of a talk move may be even more instrumental in fostering cognitive contributions (Molinari et al. 2013). Contingency means the use by teachers of student-led information or student-provided responses' to guide the flow of conversation. When a teacher requires students to provide evidence for unjustified propositions, its production is an indicator of higher-order cognitive activity (Jadallah et al. 2011).

The empirical study described in subsequent sections analyzes connections between three aspects of classroom discourse: implementation quality, critical thinking, and talk moves. Implementation quality was initially tested by RTOP scoring, to assess its association with critical thinking ([pathway of the rationale, RTOP scoring (of teachers) → implementation quality (by teachers) → critical thinking operations (by students)]). A reanalysis of implementation quality was then conducted to deepen preliminary interpretations of the relationship between implementation quality and critical thinking operations ([pathway of the rationale: RTOP scoring (of teachers) → implementation quality (by teachers) → critical thinking operations (by students) → analysis of talk moves (by teachers) → re-interpretation of critical thinking operations in terms of emerging talk patterns (by teachers and students)]).

3.2 Significance of Study and Research Questions

It should be noted that the above formulations on the relationship of discourse and cognition may be rather complicated for science teachers. As Hammerness et al. (2005) point out, teachers must recognize, make sense of, and respond to classroom events on the fly to best promote students' cognitive operations. Chan et al. (2020) proposed that understanding the multilayered dimensions of discursive interaction, and their influences on students' cognitive operations, is essential for science teachers. This teacher noticing involves teachers' pedagogic awareness of instructional actions and their influences on students. Thus, this study's outcomes can be used as a thinking tool to enhance teacher noticing, rather than relying on a sophisticated discourse-cognition model. The latter may be invisible to most science teachers, who must selectively attend to, interpret, and respond to events in a visually complex and potentially confusing classroom environment. Two research questions are addressed in this study.

1. How does implementation quality of ABI, as measured by RTOP score, stimulate students' critical thinking?
2. How do different kinds of teacher-led talk moves stimulate or inhibit students' critical thinking?

The first research question (re)tests the hypothesis that "meaningful instruction in every subject domain inherently comprises the development of critical thinking skills" (Tiruneh et al. 2018; p. 1067). In the hypothesis, meaningful instruction is emphasized, thus implicitly accepting that implementation quality is a prerequisite for predicting students' critical thinking. Through the second research question, the how aspect of the first was elucidated in detail, with the different types and proportions of enacted talk moves constituting a central component of implementation quality. Put differently, examining the frequencies of talk moves by type boosted the explanatory power of implementation quality in terms of estimating students' critical thinking.

4 Methods

4.1 Participants and Research Context

Four middle school science teachers and 92 (49 females, 43 males) seventh-grade students participated. The teachers were employed in private schools in a metropolitan city of Turkey, located in Marmara region. The students' were aged 12–13 and were of diverse socioeconomic backgrounds, some receiving free lunches. All schools had sufficient infrastructure to facilitate hands-on science inquiry activities. The teachers had different levels of experience, with one having taught science at secondary school level.

Through faculty-school cooperation, the teachers were involved in a professional development program in which their pedagogical and epistemological beliefs were explored and in-class practices enriched. This was an intense, 1-year program, in which teachers had the chance to interrogate, modify or alter their existing pedagogical theories practices. In-service training activities lasting 3 days were held on four occasions, with each incorporating pedagogical discussions. These allowed the researcher to gain insight into teachers' preconceptions of teaching science, and how these affected their classroom practices.

The teachers were engaged in ABI implementation and were introduced to student-centered science teaching activities by undertaking the student role. This was the first time the teachers had been involved in professional development centered on an ABI approach. A substantial part of the program was dedicated to preparing curriculum-based activities. The teachers worked in collaboration with the researcher(s) to transfer these activities into their own classrooms in a pedagogically appropriate way (i.e., modified for middle school students).

4.2 In-Class Implementation

4.2.1 In-Class Implementation Included Content from Three Curricular Units

Unit 1: Force and energy: (keywords: mass, weight, gravity, physical work, kinetic energy, gravitational potential energy, elastic potential energy, conservation of energy, kinetic energy loss by friction, air, and water resistance).

Unit 2: Substances and mixtures: (keywords: atom (nucleus, layer, proton, neutron, electron), property of scientific knowledge, molecule, element, chemical symbols, compound, compound formulas, homogeneous mixture, solution (solute, solvent), heterogeneous mixture, dissolution, factors affecting the rate of dissolution).

Unit 3: Interaction of light with substances: (keywords: absorption of light, objects appear black or white, color, solar energy, mirrors (flat, concave, convex), refraction of light, lenses (convex, concave), focal point).

Each ABI implementation incorporated three interwoven cycles.

Cycle 1: Initiating–Developing–Extending Teachers tried to guide students to tackle intellectual problems, with the aim of problematizing the content under consideration. Thought-provoking procedures were followed by the teachers. The main objective of the teachers was to prompt students to notice flaws in their everyday ideas and to explore the discrepancies presented to them. Once students were unsatisfied with their preconceptions, they were expected to engage in experimental processes to test their initial ideas, and to modify, enrich, or change them.

Teachers were expected to re-examine students' varying conceptual tools. Some of the students held the idea that mass and weight are the same, a belief which teachers challenged by posing discrepant questions ("If mass and weight is same, we may measure them with the same tool, with a dynamometer. However, you said that it (the tool) may only measure the weight since weight is a force. What do you think about it? Could we define the mass as a force?"). Some student-led ideas may be contrary to scientists' explanations regarding ontological commitments. When a teacher asked what happens to a salt grain placed in water, the students responded that it dissolves by going into spaces in the water. Students could then be asked ontologically thought-provoking questions such as "what is filling the spaces in water?" or "what is the stuff of the space within water before salt grains infuse into?" The students had to rearrange their ontological commitments. In addition, the epistemological underpinnings of concepts were considered. Teachers tried to problematize the darkness of empty space in the presence of numerous stars that emit light. For their part, students had to engage in epistemologically oriented reasoning into how this situation can be made knowable or researchable. In other words, in what ways and to what extent can someone elucidate the causes of empty space remaining dark even in the presence of millions of stars radiating light?

Cycle 2: Experimenting The students were directed to inspect datasets to assess the credibility of their initial hypotheses. They collected, analyzed, and interpreted data and communicated their results with others to persuade them of the explanatory power of their interpretations. The students were encouraged to differentiate sense and meaning making by engaging in experimentation to produce data-based explanations. Teachers supported students in refining their research questions, as most were unable to compose research questions by introducing significant research variables. The student groups also tended to examine a given research question by exploring similar research variable(s). Teachers therefore tried to promote variation in student-led investigations by directing them to consider alternative research variables and by supporting students to engage in multivariable reasoning. In one implementation, students decided to compare the arrival time of two objects of different weights at the bottom of an inclined plane. The teacher thus guided the students to explore mass difference, as well as the height of the plane or its inclination angle as secondary variables. In this situation, the students were encouraged to engage in multivariable experimentation. As students tended to make a limited number of experimental observations/measurements, teachers guided them to carry out multiple measurements/observations to obtain more reliable data. Student-led experiments were relatively structured, based on initial social negotiations of meaning in an initiating–developing–extending phase during which some research variables were featured while others were ignored. Although the students determined their experimental designs (constructing research questions, selecting experimental materials), teachers tried to scaffold them to bring about multivariable experimentation and reliable/valid data gathering, analysis, and interpretation.

Cycle 3: Presenting–Discussing–Finalizing In this phase, groups presented results to their classmates. The students were directed to be accountable to disciplinary norms of thinking and to interpretations obtained from their data sets. They were invited to criticize their classmates' inferences, data collection, analysis, and presentation. As the groups had addressed similar research questions, when students' own data-based interpretations were challenged by other groups' results, they sought to defend their position against counter arguments. Teachers were expected to encourage students to explain how it was possible to study the same question but

arrive at different results. At these moments, the richest verbal exchanges among the peer community were achieved. Teachers also drew attention to specific points in data presentation. For instance, the students created data tallies and graphical representations, and teachers intentionally asked why they preferred to present their data through diagrammatic representations. This attracted attention, as it was more credible to present collected data multimodally.

4.3 Data Collection

Teachers conducted eight in-class implementations. The first three were dedicated to warm-up processes in which the researcher's scaffolding was fully actualized. Data obtained from these implementations were not analyzed. In the last five implementations, teachers received no pedagogic support from the researchers, who had a passive role. All five final implementations (1693 min) were included in the analysis. The teachers conducted two implementations for the force and energy and substances and mixtures units, and one for the interaction of light with substances. Three cameras were located in the laboratories to capture all teacher–student and student–student interactions. A portable camera was used to capture one-to-one interactions. For ethical considerations, teachers and parents gave written consent to involvement in the study.

4.4 Data Analysis

The following data analysis steps were conducted.

Step 1: Verbatim transcription.

Step 2: RTOP coding to determine teachers' implementation quality levels.

Step 3: Categorizing teachers' implementation qualities and locating the hypothesis (see also first research question).

Step 4: Assigning codes to student-led remarks to identify levels of critical thinking.

Step 5: Quantifying and comparing students' critical thinking, when instructed by teachers with differing levels of implementation quality.

Step 6: Assigning codes to teachers' talk moves.

Step 7: Quantifying and comparing teachers' talk moves.

Step 8: Determining the proportions of qualitatively different talk moves that more effectively fostered critical thinking.

Gestures, mimicry, intonation, and gaze of teachers were recorded as affective dimensions of interaction (step 1). Step 2 involved RTOP coding with the aim of measuring teachers' implementation quality levels. The RTOP was used to establish average scores for implementation quality, which were then applied to predict students' critical thinking levels. In step 3, the teachers were categorized by implementation quality level and the hypothesis specified: when ABI is implemented with higher quality, it will stimulate students' higher-order critical thinking.

Martin and Hand (2009) undertook to align the components of the ABI approach with the RTOP's twenty-five item descriptors, using an interpretational analysis approach to select the latter. Table 1 displays the equivalency established between ABI components and RTOP items. Thirteen of the twenty-five items were consolidated into four categories (student voice

(five items), elements of argument (five items), teacher's role (two items), and questioning (one item)).

The RTOP provides four intervals (0–4) to describe implementation quality: “0–1” low level; “1–2” acceptable; “2–3” has merit; and “3–4” high level. Three coders, as well as the researcher, conducted RTOP scoring. For category 1, student voice and category 3, science argument, a notably lower intercoder reliability coefficient was observed (0.78). Through constant discussion of divergent results, this coefficient was increased to 0.91.

In step 4, the critical thinking operations of students were analyzed. Codes were assigned to student-led remarks to determine whether levels of critical thinking fluctuated from one implementer (teacher) to another. Results of three seminal research studies were used to compose a catalogue for analyzing critical thinking: (a) Facione (1990): e.g., interpretation, analysis, evaluation, inference, explanation, (b) Ennis (2011): e.g., basic clarification, inference, advanced clarification, (c) Grimberg and Hand (2009): e.g., observation, comparison, analogy, clarification, investigation design.

Grimberg and Hand's (2009) indicators were most influential, as their study concentrated on the ABI approach. Fourteen descriptors of critical thinking were divided into three (hierarchical) themes: perception (lower level), conception (moderate), and abstraction (higher level) (Table 2). Three researchers assigned codes from the catalogue to classify student-led critical thinking operations. Although coders had difficulty differentiating higher-level critical thinking operations, such as advanced clarification and inductive reasoning, intercoder agreement was sufficiently high (89%). After coding all relevant student-led remarks, these were counted and compared in step 5. The objective was to ascertain in which teachers' classrooms students attained more higher-order critical thinking levels. This allowed implementation quality to be associated with variation in critical thinking.

To enrich the preliminary RTOP-based interpretations of implementation quality, teacher-led talk moves were also analyzed. The RTOP-based interpretations provided an initial understanding of the relationship between implementation quality and critical thinking level (steps 1 to 5). Additional analysis of talk moves allowed this relationship to be comprehended much more clearly. This fine-grained reanalysis demonstrated that critical thinking was stimulated not only by type of teaching activity but also by type of talk move.

In step 6, observed talk actions (remarks) of teachers were assigned codes from the coding catalogue (Table 3) developed for this study. The catalogue was constructed, based on the findings of previous studies, and on the nature and structure of spoken language in the implementations (Mercer 2010). It incorporates seven higher-order themes and several characterizing codes to systematically classify teacher-led remarks and includes both authoritatively oriented and dialogically oriented moves. Three experts conducted coding procedures, achieving an intercoder reliability over 0.88. The coders acted as external auditors and consultants, examining both the process and the product of the method to assess their accuracy. In step 7, extracted types of talk move were counted. This permitted detection of whether teachers enacted talk moves in different proportions, and was useful in determining which qualitatively different talk moves were most able to foster critical thinking (step 8).

Table 1 Associations between the RTOP* and argument-based inquiry approach's quality indicators

The quality indicators of the RTOP	The quality indicators of the argument-based inquiry approach
#Theme 1: Student voice	
1. Instructional strategies respected students' prior knowledge/preconceptions.	Connections: There is an emphasis on determining student knowledge and building teacher plans based on this knowledge.
2. Focus and direction of lesson determined by ideas from students.	Connections: Teacher builds or activates students' prior knowledge with some evidence of using it to make instructional decisions.
3. Students communicated their ideas to others.	Focus on learning: Student sharing with argumentation/connections in either small group, group to group or whole group. Connections: Language activities flow naturally throughout the ABI. Science argument: Teacher promotes linkages to big ideas and begins to promote debate on these ideas.
4. High proportion of student talk and a significant amount was student to student.	Focus on learning: Student sharing with argumentation/connections in either small group, group to group or whole group. Dialogical interaction: Communication effectively varies from teacher to student and from student to student according to the situation.
5. Students' questions and comments determined focus and direction of classroom discourse.	Connections: Teacher effectively builds or activates student prior knowledge with evidence of using this to make instructional decisions. Dialogical interaction: Teacher is not compelled to give right answer shifting focus to the big idea. Teacher uses all levels of questioning and adjusts levels to individual students.
#Theme 2: Teacher role	
1. Teacher acted as resource person, supporting and enhancing student investigations.	Focus on learning: Teacher effectively plans for teacher and student instruction as needed and appropriate.
2. The metaphor "teacher as listener" was very characteristic of this classroom.	Dialogical interaction: Teacher used questions to explore student thinking. Teacher's response to student answers is probing, connects, and extends, questions.
#Theme 3: Science argument	
1. Students were actively engaged in thought provoking activities that involved critical assessment of procedures.	Connections: Science activities promote big ideas clearly and extend students learning. Connections can be seen from beginning to end and are articulated by students.
2. Students were reflective about their learning.	Science argumentation: Teacher demands connections between question, claims, evidence and reflection.
3. Intellectual rigor, constructive criticism, and the challenging of ideas was valued.	Focus on learning: Student sharing with argumentation/connection in small groups, group to group and whole group with few prompts. Science argumentation: Teacher promotes linkage to big ideas and promotes debate on these ideas.
4. Active participation was encouraged and valued.	Science argument: Teacher requires students to link claims and evidence. Teacher scaffolds questions, claims, evidence and reflection. Teacher promotes linkages to big ideas and debate of these ideas.
5. Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	Science argumentation: Teacher scaffolds questions, claims, evidence and reflection. Teacher promotes linkages to big ideas and debate of these ideas.
#Theme-4: Questioning	
1. Teacher questioning triggered divergent modes of thinking.	Dialogical interaction: Students are asked to explain and challenge each other's responses rather than the

Table 1 (continued)

The quality indicators of the RTOP	The quality indicators of the argument-based inquiry approach
	teacher passing judgment. Teacher asks many layered questions. Teacher is not compelled to give right answer shifting focus to the big idea.

*Adapted from Martin and Hand (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, 39, 17–38

5 Findings

Results of the RTOP scoring are presented first, followed by a comparison of critical thinking levels by students in each teacher's classroom. These two results are then combined to determine whether the implementation quality of an immersive instructional approach

Table 2 Levels of critical thinking operations of students

Operation	Indicators
Perception zone: lower level critical thinking	
Presenting relevant experiences (L*)	Stating and sharing individual-based experiences, externalizing subjective experiences
Differentiating (L)	Referencing to common or different characteristics, states of two or more claims, comparing conclusions, predictions, matching or contrasting opinions, conclusions, predictions, etc., comparing prior and recent thinking, sayings, and opinions.
Conception zone: moderate critical thinking	
Illustrating (M**)	Introducing relevant examples, instances, samples, trials of events, concepts, etc.
Analogizing (M)	Mapping elements from a source domain (well-understood situation) into a target domain (nonfamiliar situation)
Interpreting 1 (basic, low level) (M)	Simple clarification for making shared utterances common, pieces of knowledge and information that stimulate clarification supporting other explanations, sayings, arguments, assumptions, opinions, positions, ideas, etc., decoding significance, simply clarifying meaning within discourse, simply uttering hypothetical opinions
Inferencing (low level induction) (M)	Making projections, proposing guesses
Concluding (M)	Presenting conclusive claims, summarizing concluding remarks
Interpreting 2 (advanced, high level) (M)	Define, redefine terms, concepts, definitions by using appropriate criteria, examples, instances, samples, making operational definitions to clarify the meaning, arguments, sayings, claims, assertions, etc.
Causal reasoning (M)	Identifying of a cause and its effect, defining a simple causation between two or more properties of an object, instance, event, sample, etc.
Abstraction zone: higher level critical thinking	
Elucidating (H***)	Offering unproved scientific inference (by means of scientific principles, theories, laws, etc.), justifying thinking procedures
Reflecting (H)	Assessing other's claims, judging the credibility of an information source, examining other's ideas, judging other's deduction, inductions, definitions
Deductive reasoning (H)	Reasoning that links general premises to a specific one, conclusion is based on the concordance of multiple premises that are generally assumed to be true
Inductive reasoning (H)	Reasoning that links few examples to general premises, making material and generalized inferences (roughly "induction"),
Creating proposals (H)	Revising or readjusting classmates' ideas, stating new ways of thinking and talking

L*Lower level critical thinking; M**Moderate critical thinking; H***Higher level critical thinking

Table 3 Detailed list and description of enacted talk moves during argument-based inquiry implementations

Higher-order categories	Subcategories as TDMs	Description
Evaluating	Presenting logical expositions	Teacher provides canonical science knowledge as in the form of science communities' logical propositions.
	Direct affirmation	Teacher evaluates and determines the credibility of an uttered response by considering his/her teaching agenda's content and flow.
	Rejecting	Teacher declines a student-led response that is out of the scope or context of the discussion
	Softly rejecting	Teacher does not directly or explicitly reject the response, however, implies that the response may not be welcomed.
	Affirmation-cum-direct-instruction	Teacher welcomes a student-led response that is instrumental for his/her teaching agenda, then, gives additional detailed information regarding given response's content.
	Verbal cloze	Teacher stops his/her talking and invites the students to fill the gap within the utterance led by him/her.
Communicating	Probing	Teacher presses students to probe, elicit and externalize their responses' underlying meanings.
	Requesting for clarification	Teacher promotes students to make their utterances clearer or more intelligible to the peer community.
	Reformulating	Teacher revoice a student-led response to make it more comprehensible for all members of the community.
Monitoring	Focusing	Teacher grasps students' attention to a particular response or a focal point that is important for the progression of classroom discourse.
	Summarizing (consolidating)	Teacher captures and summarizes student-led responses and mentions the consolidated ideas emerged in students' talks by categorizing.
	Selecting and eliminating	Teacher selects (highlights, features) or ignores (eliminates) specific responses and implies the direction of classroom discourse by emphasizing a being selected student-led response.
	Asking about mind change	Teacher stimulates students to check whether their ideas are revised, modified, expanded or totally altered during discussions.
Legitimizing	Asking for evaluation (student-led)	Teacher prompts students to make an evaluation, assessment and criticism of a provided response to foster student-student verbal exchanges.
	Asking for evaluation (case-based)	Teacher presents a case, demonstration or thought-provoking idea for the evaluation and judgement of students.
	Asking for evaluation (teacher-led)	Teacher externalizes his/her meaning position for the critical assessment or judgement of students.
Challenging	Conceptual tools	

Table 3 (continued)

Higher-order categories	Subcategories as TDMs	Description
Evidencing	Ontological commitments	Teacher challenges student-led ideas by referring to their conceptual flows by playing devil's advocate role. Teacher contradicts student-led propositions regarding their ontological fallacy to foster further thinking.
	Epistemological underpinnings	Teacher interrogates confusions of students by referring to more or less dependable ways of knowing.
	Asking for supported reasoning	Teacher asks students to warrant their claims by referring to data or relevant examples as in the form of evidences.
	Praising use of evidence	Teacher reinforces students who employ evidence while asserting their meaning positions.
	Referring in-text information (evidence)	Teacher directs students to attribute to text-based information as in the form of evidence for their meaning positions.
Observe–compare–predict	Observing	Teacher requires students to make observations or present their observation-based experiences to make sense of a proposed opinion.
	Comparing	Teacher guides students to make comparisons of ideas, arguments, cases, and etc.
	Predicting	Teacher prompts students to make projections regarding a science phenomenon through conjectural or hypothetical reasoning.

stimulates students' critical thinking. Finally, selected episodes from in-class implementations are interpreted to show how their quality, types of talk moves, and students' critical thinking interact. For these purposes, statistical values of observational data are expressed as percentages and standard deviation as *z*-scores. *Z*-scores were calculated for each variable (implementation quality, talk move, critical thinking) by taking observed values, subtracting the mean of all observations, and dividing the result by the standard deviation of all observations. As Field (2013, p. 26) stressed that *z*-score is highly significant in interpreting trend lines between two or more means obtained from observational data. Numeric distributions calculated for implementation quality level, talk move typology, and critical thinking level were converted into *z*-scores as a new distribution with a mean of 0 and a standard deviation of 1.

5.1 How Does Implementation Quality of the ABI Approach Stimulate Students' Critical Thinking?

Average implementation quality scores as measured by the RTOP (Fig. 1) show that teachers C (mean = 2.92; *z*-score = + 0.35) and D (mean = 3.22; *z*-score = + 0.80) were better implementers compared with the other two teachers. Teachers A (mean = 1.68) and B (mean = 1.75) achieved lower implementation quality scores. The *z*-scores show that, while teacher A (*z*-score = − 1.49; more than 1 standard deviation below the mean implementation quality) and teacher B (*z*-score = − 1.38) tried implement the ABI approach, their in-class practices did not adequately serve the purposes of the approach. By inference, teachers C and D seemed to implement the ABI approach (nearly) exactly as

defined or were better able to carry out its core components compared with other two teachers in the context of the study.

Low-level critical thinking was more common among students in teacher A's (22.6%) and teacher B's (23.3%), as compared with teacher C (18.6%), and especially teacher D (9.2%) (see also Fig. 2). This difference was significant between teacher D (z -score = -1.1 ; more than 1 standard deviation below the mean calculated for low-level operations) and teachers A and B (z -scores = $+1.02$ and $+1.13$ respectively; both more than 1 standard deviation above the mean calculated for low-level operations).

A similar pattern was also evident for moderate level critical thinking operations. The students' critical thinking was predominantly at a moderate level in all four classrooms (teacher A: 68.8%; B: 59.7%; C: 57.9%; D: 55.4%). However, z -scores showed that particularly in teacher A's classroom (z -score = $+2.57$; a value outside ± 1.96 is significant at $p < 0.05$), the students pervasively generated remarks at a moderate critical thinking level compared with teacher C's (z -score = $+0.11$) and teacher D's (z -score = $+0.72$) students.

Significant differences were also observed in the generation of higher-order critical thinking across classrooms (Teacher A: 8.6%; B: 17%; C: 23.5%; D: 35.4%). The difference in students' higher-order critical thinking operations was most apparent between teacher A (z -score = -1.75 ; more than 1 standard deviation below the mean calculated for higher-order operations) and teacher D (z -score = $+1.03$; more than 1 standard deviation above the mean calculated for higher-order operations). It was thus concluded that, as teachers became better implementers of the approach, students' achieved higher-order levels of critical thinking.

5.2 Types of Teacher-Led Talk Moves and Students' Critical Thinking Operations

5.2.1 Evaluating Moves

One of the observed talk moves was categorized as evaluating, as exemplified in Table 4. Teacher A talked with students about molecules and compounds, initiating a classroom

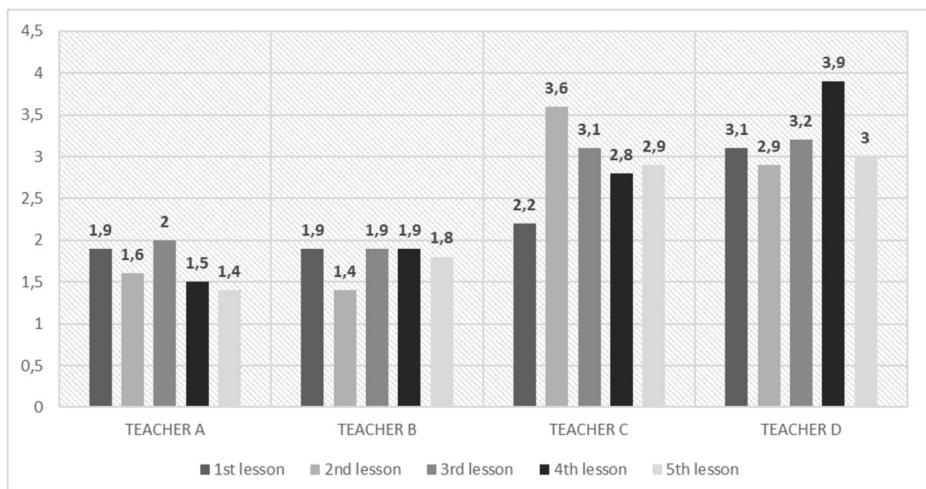


Fig. 1 The teacher's RTOP scores indicating their implementation quality

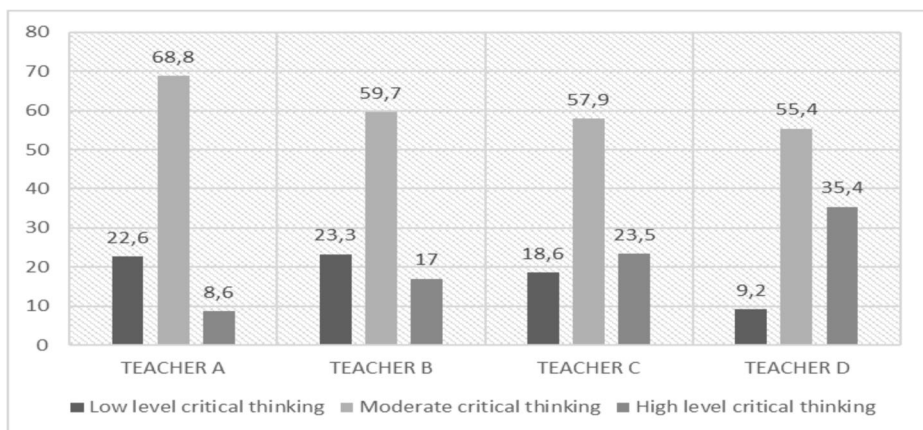


Fig. 2 The students' critical thinking operations' levels in different teachers' classroom

discourse (line 1) through a question expected to direct students to recall previous content. S4 (line 2) provided an immediate response, referring to the key concepts handled in the previous lesson. Teacher A posed an additional question to elicit from students the definition of “compound.” S3 (line 4) uttered an ill-structured response, which the teacher implicitly indicated was not acceptable before posing a further question (line 5). The teacher seemed to adhere to a strict instructional agenda, and did not use student-led information to unfold dialogue. Teacher A tried to find a more rational answer to the question (line 7), and so rehearsed the query by playing a conventional classroom game: know the thing that is in my head. Teacher A mostly delivered technical explanations to students via confirmatory talks. Once the teacher had enacted this group of moves overly often, the students seemed to resort to baseless responses evaluated at a low level of critical thinking.

All teachers enacted evaluating moves during the implementations (Fig. 3), but at varying frequencies. Both teacher A (21.6%; z -score = +1.58, more than 1 standard deviation above the mean calculated for evaluating moves) and teacher B (19.8%; z -score = +1.26) relied more on evaluating moves compared with teachers C (11.5%; z -score = -0.17) and D (7.4%; z -score = -0.89; nearly 1 standard deviation below the mean calculated for evaluating moves). Frequent use of evaluating moves may have reduced student-led voices or their speaking time in the classrooms of teachers A and B. This might diminish the number of discursive opportunities for students to attain higher-order critical thinking.

5.2.2 Legitimizing Moves

Legitimizing moves were used by teachers to stimulate students to criticize their classmates' opinions. In the subtopical episode described in Table 5, the group negotiated which variables (inclination angle, weight of objects, skin friction) influence an object's motion on an inclined plane. The students were about to design and conduct their experiments by considering the variables discussed in this section.

Teacher D (line 1) invited students to draw conclusions about the factors influencing an object's motion on an inclined plane. S11 (line 2) claimed that weight (“The loaded truck arrives immediately...”) is an important factor affecting the arrival time of an object at the bottom of the plane. Teacher D maintained a neutral position, avoiding a rapid evaluation of

Table 4 The teacher A displayed some of the evaluating moves during discussions about substances and particles

Turn	Speaker	Utterance	Connection to the quality indicator	Type of talk move	Critical thinking operation
1	T	Yeah, what did we talk about last lesson?		The teacher initiated the dialogue.	–
2	S4	Molecule, compound. And... Atom!		–	Presenting relevant experiences
3	T	What is the compound, for instance?	The metaphor “teacher as listener” may be a characteristic of this classroom. (Table 1; theme 2: teacher role, item 2)	Probing	–
4	S3	Some things in the air. For example, oxygen.		–	Illustrating
5	T	Is oxygen a compound or not?	Focus and direction of lesson is not determined by ideas from students. (Table 1; theme 1: student voice, item 2)	Comparing	–
6	S3	Is not it?		–	None
7	T	I do not think so. // Is oxygen a molecule or a compound?	Students’ questions and comments do not determine focus and direction of classroom discourse. (Table 1; theme 1: student voice, item 5)	Rejecting // Comparing	–

the answer. The teacher (line 3) then referred to the previous conversations and invited Linda to evaluate S11’s assertion, in light of her contrary position. Linda was thus assigned as an epistemic authority, with the right to legitimate an answer in terms of its plausibility. It can be

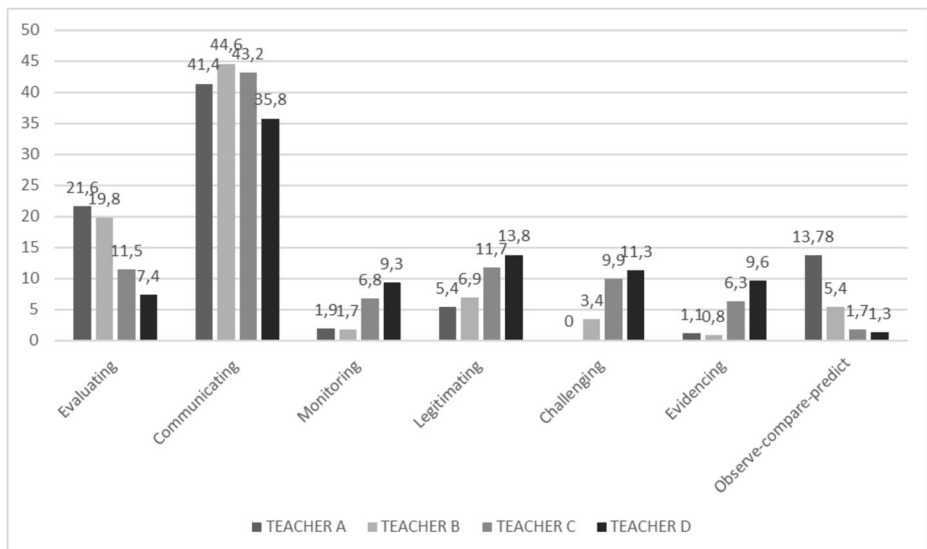
**Fig. 3** Observed types and percentages of teacher-led talk moves

Table 5 Teacher D enacted legitimating moves in the discussions about force and motion

Turn	Speaker	Utterance	Connection to the quality indicator	Type of talk move	Critical thinking operation
1	T	So, let us come to a conclusion now. Which of these vehicles first arrive at the end of the slope?		The teacher initiated a conclusive episode.	—
2	S11	The loaded truck arrives immediately. It cannot even stop. Very heavy load.		—	Inferencing
3	T	Yeah, Wendy, that's different than you said. What do you think of that?	Students' questions and comments determine focus and direction of classroom discourse. (Table 1; theme-1: student voice, item 5) Students are asked to explain and challenge each other's responses rather than the teacher passing judgment. (Table 1; theme 4: questioning, item 1)	Asking for evaluation (student-led)	—
4	S14	But it may not be. Now I think the truck has too much load. But if this truck gets heavier, its wheels will stick to the ground. Hard to move. Crawling wheels everywhere. Then it will slow down.		—	Causal reasoning
5	T	Yes, Ahmet, your friend agrees with you at some point. But she said it slows down when its wheels touch the ground. // Yes, class and Ahmet, do you have an answer to that?	Intellectual rigor, constructive criticism, and the challenging of ideas are featured. (Table 1; theme-3: science argument, item 3) Students are asked to explain and challenge each other's responses rather than the teacher passing judgment. (Table 1; theme 4: questioning, item 1)	Summarizing // Asking for evaluation (student-led)	—
6	S11	No matter how hard the wheels stuck, that truck cannot stop on a steep hill, and it goes quickly. I mean, it just too heavy. Even though its wheels rub on the ground, it is pulled down.		—	Inferencing
7	T	OK, but your friend is not convinced. Is there an example to convince him?	Teacher requires students to link claims and evidence. (Table 1; theme 3: science argument, item 4)	Asking for supported reasoning	—
8	S11	For example, let us tie them both up and leave. No rope can hold a heavy truck. It cannot hold the heavier one even its wheels		—	Creating proposals

Table 5 (continued)

Turn	Speaker	Utterance	Connection to the quality indicator	Type of talk move	Critical thinking operation
		stuck to the ground. The rope is immediately stretched and ruptured.			

inferred that Linda's comments determined the focus and direction of the dialogue. Teacher D tended to ask students to generate alternative points of view on the motion of objects on an inclined plane, frequently soliciting evaluation moves. S14 (Linda, line 4) responded to her classmates by noting that a heavier object will have greater contact friction, slowing its movement along the inclined plane.

Teacher D (line 5) summarized Linda's point and asked S11 to react to her reasoning, which consisted of an alternative explanation for the behavior of objects on an inclined plane. Teacher D seemed to follow a discursive path which featured intellectual rigor, constructive criticism, and challenging of ideas. Through frequently enacted legitimating moves, students in teacher D's classroom had to take their peers' claims seriously and evaluate their credibility. In line 6, S11 provided his inference in response to Linda, but the teacher (line 7) immediately requested evidence-based reasoning to make S11's proposition more persuasive. S11 (line 8) then provided an enriched example to convince Linda that the weight of an object may overcome the force of friction while on an inclined plane. S11's well-structured response was evaluated as a research proposal (higher-order critical thinking operation; Table 2, revising or readjusting classmates' ideas, stating new ways of thinking and talking) to understand the principles of an inclined plane as a simple machine.

This subtopical episode showed that frequent and timely use of legitimating moves may be positively correlated (see Table 5) with the quality descriptors of ABI, and in turn, with critical thinking by students. Teacher C (11.7%; z -score = +0.22) and particularly Teacher D (13.8%; z -score = +0.86; nearly 1 standard deviation above the mean calculated for legitimating moves) tended to employ legitimating moves more frequently than teachers A (5.4%; z -score = -1.68; more than 1 standard deviation below the mean calculated for legitimating moves) and B (6.9%; z -score = -1.22). Frequent use of legitimating moves seemed to stimulate students' critical thinking operations, particularly in the realm of abstraction (see S11's remark in line 8, Table 5).

5.2.3 Challenging Moves

Teachers were expected to persuade students of the existence of alternative explanation systems that may be more illuminating than their own, everyday accounts of scientific phenomena. This seemed to be facilitated by teachers performing challenging moves that made fallacies in student-led reasoning explicit. Once students were challenged by the teacher-led alternative thinking, they became dissatisfied and sought to modify, advance, or alter their existing mental models. The purpose of a challenging move was not to falsify a novice individual mercilessly. It was intended to guide students to notice that their own explanations may be less adequate in making sense of a phenomenon, and should therefore be revised, enriched, or changed.

Considering all previous conversations, S8 (line 1) cited the size difference between salt and stone to explain why the former dissolves in water but the latter does not (Table 6). The stone is large and the salt small, according to S8. The idea that stone cannot get between water molecules, whereas salt can and is thus dissolved, was a proposition that S8 could not explain but believed.

Teacher D (Table 6; line 2) injected alternative (challenging) reasoning into the classroom discourse by guiding students to make an instant observation. Teacher D broke the stone into smaller pieces, making it comparable with tiny salt particles, in an effort to create the experimental conditions to test S8's proposition regarding size difference. The teacher allowed students to make a purposeful observation to test whether one of their assertions worked in real conditions, and might therefore be valid. If S8's argument was plausible in theory, it had to be

Table 6 Teacher D performed challenging moves in the properties of solutions discussion

Turn	Speaker	Utterance	Connection to the quality indicator	Type of talk move	Critical thinking operation
1	S8	But there is a size difference. Salt consists of small solids. But the stone in your hand is big. Where will it go in the water? Salt goes in.		—	Interpreting 2 (advanced, high level)
2	T	OK. Let me crumble this stone. I will make it like sand grains and throw it in... Then will it dissolve? Is it soluble then? How about that?	Students' questions and comments determine focus and direction of classroom discourse. (Table 1; theme 1: student voice, item 5)	Asking for evaluation (case--based)	—
3	S8	No, it is insoluble.		—	Inferencing (low-level induction)
4	T	Why not dissolve... Best to try... Look, I even made this stone as small as powdered sugar. Is it not dissolved then?	Intellectual rigor, constructive criticism and the challenging of ideas are featured. (Table 1; theme 3: science argument, item 3)	Challenging	—
5	S12	No. How could it be solved?		—	Interpreting-1 (basic, low level)
6	T	Why not dissolve? Both are tiny, huh? It is now as small as then salt grain, even smaller than it as you insisted on from the beginning of the lesson. I think this idea was always dominant. Then it must logically dissolve.	Students were actively engaged in thought provoking activities that involved critical assessment of procedures. (Table 1; theme 3: science argument, item 1)	Challenging	—
7	S12	It is insoluble! This is like olive oil and water. Because normally these have to go into each other. It should look like a single matter. But here it is not. Because of the type of...		—	Inductive reasoning

tested in practice. In terms of the approach's quality indicators, Teacher D used student-led information (asking for evaluation (case-based)) to determine the focus and direction of classroom dialogue.

S8 (line 3) seemed to renounce his idea. Teacher D (line 4) asked again whether stone particles will remain undissolved if the stone is broken into ever smaller fragments like powdered sugar. The basic instructional purpose of the teacher was to prompt students to test the idea of whether a piece of (insoluble) stone will form a solution. Students seem quite conflicted by this idea. In this intellectually rigorous learning environment, students had to interrogate their own theories regarding solvation. S12 (line 5) expressed the impossibility of the idea but was unable to develop a satisfactory clarification.

Teacher D challenged students to clarify the reason(s) for the insolubility of stone by considering microscopic rather than macroscopic properties. Following S12's response (line 5), the teacher (line 6) stressed that size difference was the dominant idea. Through teacher D's questioning (line 6), students had the opportunity reconsider their reasoning on insolubility and noticed that their ideas had changed over the course of the discussion talks. However, as the teacher emphasized (line 6), internal consistency during idea sharing and defending would be preferable. Teacher D behaved as a rigid discussant, debater, and negotiator, whose students seemed to engage in thought-provoking cognitive activities involving frequent critical assessment of propositions (connecting with quality indicators). Based on this, S12 (line 7) tried to resolve the conflict through an olive oil–water analogy. S12 referred to the term homogeneity, and implied that type of substance should be considered responsible in solvation instead of grain size. S12's response (line 7) was evaluated as a higher-order critical thinking operation since she proposed a more generalized explanation for solvation, in response to the teacher's discrepant questioning.

Challenging moves seemed to prompt students to engage in higher-order critical thinking. Teacher A did not stage any challenging moves (z -score = -1.93 ; nearly 2 standard deviations below the mean calculated for challenging moves), while teacher B displayed them rarely (3.4%; z -score = -1.12 ; more than 1 standard deviation below the mean calculated for challenging moves) (Fig. 3). Teachers C (9.9%; z -score = $+0.4$) and D (11.3%; z -score = $+0.74$) enacted challenging moves with considerably higher frequencies than teachers A and B. When student-led ideas were consistently questioned by teacher C and particularly teacher D, the students had to improve their clarifications, scrutinize their ideas in light of counter arguments, and create new ways of thinking to respond to the discussant teachers.

5.2.4 Monitoring Moves

Monitoring moves were enacted to involve students in monitoring streams of conversation, allowing teachers and students to pursue goals based on identical implicit time frames. Monitoring moves seemed crucial to ensuring conceptual synchronization between teachers and students. In the subtopical episode (Table 7), teacher C expected students to establish relationships between concepts of heat, light, and energy.

The main topic of discussion was whether heat and light, different manifestations of energy, can be sources of each other and in which conditions this is possible. Teacher C noticed confusion on the part of the students in connecting concepts of heat, light, and energy. The teacher therefore pursued a specific student-led example to reinstate discussion. Several examples had been given by the students, one of which teacher C selected (Table 7, line 1; "Shall we talk that point?") as the most appropriate to advance classroom dialogue. Teacher C seemed to feature one student-led example (e.g., Table 7, line 1: "How does the sun warm us

Table 7 The teacher C staged monitoring moves when the students compared light, heat, and energy concepts

Turn	Speaker	Utterance	Connection to the quality indicator	Type of talk move	Critical thinking operation
1	T	But now it is too confused. How does the sun warm us and other planets? Shall we talk that point?	Students' questions and comments determined focus and direction of classroom discourse. (Table 1; theme 1: students voice, item 5)	Selecting and eliminating	—
2	S1	When the light comes to my hand, for example, my hand is warming up. Warms up with light.		—	Illustrating
3	T	So, the sun is a source of heat?	Teacher's response to student answers is probing, connecting and extending. (Table 1; theme 2: teacher role, item 2)	Reformulating	—
4	S1	Yes.	Teacher's response to student answers is probing, connecting and extending. (Table 1; theme 2: teacher role, item 2)	—	None
5	T	Is heat and light the same?		Probing	—
6	S7	No. They are different.	Students are asked to explain and challenge each other's responses rather than the teacher passing judgment. (Table 1; theme 4: questioning, item 2)	—	Differentiating
7	T	But your friend said he thought that they are the same.		Focusing	—
8	S7	But not! I disagree.	Teacher's response to student answers is probing, connecting and extending. (Table 1; theme-2: teacher role, item 2)	—	None
9	T	Why not? // Your friends said that heat and light can be the same.		Probing // Focusing	—
10	S7	Then every warming object comes out light. For example, my hand is warm. Is the light coming out of my hands now? For example, in the morning teapot boils, light does not come out.		—	Reflecting

and other planets?") while eliminating others from the conversation to provide a clearer pathway for students to ponder rather sophisticated science concepts.

The teachers mostly enacted focusing submoves. For example, in line 7, teacher C invited S7 to notice that one of his classmates had an alternative idea to his ("But your friend said he thought that..."). At this point, teacher C tried to direct S7's attention to an alternative point of view through a focusing move. The teacher's aim was to notify the student(s) regarding what is going on in the discussion. Teacher C (and others) referred to respondents' remarks by quoting them (line 9: "Your friends said that..."). When teacher C frequently injected this request into

the discussion (line 7, line 9), S7 was cognitively obliged to elaborate on and connect his assertions with others. This resulted in higher-order critical thinking by S7 (line 10) as he was called upon to judge the credibility of a classmate's idea.

Teachers D (9.3%; z -score = +0.91; nearly 1 standard deviation above the mean calculated for monitoring moves) and C (6.8%; z -score = +0.15) performed monitoring moves more frequently than teachers A (1.9%; z -score = -1.32; more than 1 standard deviation below the mean calculated for monitoring moves) and B (1.7%; z -score = -1.38) (Fig. 3). Monitoring moves appeared to stimulate more sophisticated forms of student-led critical thinking. For instance, when teachers frequently used monitoring moves, students had more opportunity to apply inductive or deductive reasoning by building on others' ideas (Table 7; S7, line 10).

5.2.5 Evidencing Moves

Particularly during presenting–discussing–finalizing phases, students were invited to present their data-based interpretations. Through evidencing moves, they were encouraged to employ supported reasoning (Table 5, line 7; “Is there an example to convince him?”). With respect to quality indicators, evidencing moves seemed functional in demanding that students link claims with evidence. There were few traces of evidencing moves in the implementations of teachers A (1.1%; z -score = -1.28) and B (0.8%; z -score = -1.36; more than 1 standard deviation below the mean calculated for evidencing moves) compared with teachers C (6.3%; z -score = +0.07) and D (9.6%; z -score = +0.94; nearly 1 standard deviation above the mean calculated for evidencing moves). Evidencing moves seemed to promote higher-order critical thinking operations among students (see also Table 5, S11: line 8).

5.2.6 Observe–Compare–Predict Moves

The teachers guided students to perform basic science skills: observing, comparing, predicting. Students were asked to make observations of a science demonstration held in the classroom or to recall an individual experience in the form of an observation. Teachers then called upon students to compare events, ideas, arguments, cases, etc. (“Is oxygen a compound or not?”). In this situation, students tended to rely on their prior knowledge and put forth simplified comparisons. This implies that frequent requests to make comparisons demanded less cognitive effort on the part of students, thus reducing higher-order critical thinking. Teachers also encouraged students to make estimations, asking them to evaluate the possibility of an idea (“Can something like that happen?”).

This set of moves was observed more in the classrooms of teachers A (13.78%; z -score = +2.59; more than 2 standard deviations above the mean calculated for observe–compare–predict moves; a value outside ± 1.96 is significant at $p < 0.05$) and B (5.4%; z -score = +0.48) than those of teachers C (1.7%; z -score = -0.45) and D (1.3%; z -score = -0.55) (Fig. 3). Moves asking for observation and comparison seemed to stimulate students' critical thinking, particularly in the realm of perception. As seen in Fig. 2, in the classrooms of teachers A (perception = 22.6%; conception = 68.8%) and B (perception = 23.3%; conception = 59.7%), more student-led remarks were pitched at the level of perception and conception compared with teachers C (perception = 18.6%; conception = 57.9%) and D (perception = 9.2%; conception = 35.4%). This can be explained by taking the nature of observe–compare–predict moves into account. In teacher D's classroom, occurrences of critical thinking in the realm of perception (9.2%) were dramatically lower when compared with the other teachers. This

implies that using observe–compare–predict moves frequently in the absence of other talk moves, such as challenging, legitimating, or monitoring, may not be instrumental in stimulating higher-order critical thinking.

5.2.7 Communicating Moves

Communicating moves (probing, requesting clarification, reformulating) were used extensively to help students express their positions more clearly. The teachers used probing moves to elicit the underlying meaning of a given response (“Why did you think so?”), and requested clarification if a student’s comment was not intelligible (“What do you mean by this?”). Reformulating moves were enacted to paraphrase or transform students’ remarks, and make them more comprehensible (“So, you’re saying if the truck is loaded or not, then is the steepness of the slope important?”). The teachers performed communicating moves to ensure that all participants in classroom discussions had an understanding of others’ contributions. Communicating moves can be attached to quality indicators in terms of their instrumentality in probing, connecting, and extending students’ opinions (see Table 7 for connections to quality indicators).

It was observed that communicating moves mostly stimulated students’ critical thinking in the realm of conception (illustrating, analogizing, interpreting 1 (basic, low level), interpreting 2 (advanced, high level)). Teacher A (41.4%; z -score = + 0.26) and particularly teacher B (44.6%; z -score = + 1.06; more than 1 standard deviation above the mean calculated for communicating moves) displayed communicating moves considerably more frequently than teacher D (35.8%; z -score = − 1.13; more than 1 standard deviation below the mean calculated for communicating moves) (Fig. 3). This may explain why teacher A’s (68.8%) and teacher B’s (59.7%) students’ cognitive operations remained mostly at the conception level compared with teacher D (55.4%) (Fig. 2). By inference, while communicating moves are needed for higher quality ABI implementations, they are not sufficient to stimulate higher-order critical thinking operations.

To sum up, the ABI approach may enhance students’ higher-order critical thinking when implemented with higher quality. Furthermore, there is a relationship between implementation quality, different typologies of talk moves, and critical thinking. Some talk moves (challenging, legitimating, monitoring, evidencing) were found to stimulate students’ higher-order critical thinking very effectively. Other talk moves (knowledge providing, evaluating, observe–compare–predict), proved less able to encourage students to apply higher-order critical thinking. Importantly, communicating moves seemed to be instrumental in triggering and sustaining elicited verbal interactions, both among students and between students and teachers. These moves appeared to promote higher-order critical thinking, but their effect was most apparent when used alongside the other, more stimulating talk moves listed above.

6 Discussion

6.1 Relationships Between Implementation Quality and Critical Thinking Operations

Previous studies reported that when ABI is conducted with higher quality, the variance in conceptual understanding between high-achieving and low-achieving students is reduced (e.g.,

Akkus et al. 2007; Poock et al. 2007). This study did not aim to explore the effects of ABI on student-led conceptual understanding. Nonetheless, like the studies above, it found that an ABI approach may also be considered an effective instructional tool for stimulating students' critical thinking when conducted with higher quality. Other, similar ways of evoking students' critical thinking also exist: problem-based learning, scenario-based learning, hands-on learning, case-based learning or case study method, and Socratic questioning (Phillips and Mackintosh 2011). As shown here, ABI can be appended to the list of techniques which stimulate critical thinking. The contribution of the present study is that implementing an immersive approach may not be enough to induce higher-order critical thinking. It is also necessary for an immersive approach to be implemented with higher quality, or exactly as defined, to authentically elicit higher-order critical thinking.

There is an ongoing debate over whether a generalist (direct teaching of critical thinking as a separate course) or an immersive approach (teaching critical thinking in an embedded manner, e.g., the ABI approach) is more effective in arousing critical thinking. Some studies support teaching critical thinking directly in the context of a designated course (Ennis 1962; Tiruneh et al. 2018), thus cultivating excellence in executing the basic operations of critical thinking (Ennis 1985; Tiruneh et al. 2017). Others argue that critical thinking should be experienced by students as a matter of course within relevant and supportive learning environments in which subject matter instruction is prominent (McPeck 1981; Sternberg and Williams 2002). This study confirmed the importance not just of particular approaches to promoting critical thinking but of the quality of their implementation as characterized by specific teacher-led talk moves. Contrasting previous studies, it crystallized the relationship between implementation quality and talk moves. More specifically, the following practices were found to increase implementation quality and promote students' critical thinking:

- Teachers had to give greater prominence to students' voices in the classroom (communicating moves, evaluating moves).
- There had to be vigorous debate among the peer community (legitimizing moves, evidencing moves, challenging moves).
- Academic rigor and reflective discourse had to be adopted by both the teacher and the peer community (legitimizing moves, evidencing moves, challenging moves).
- Students had to consciously maintain conceptual and procedural awareness of classroom happenings (monitoring moves).

These discursive features were also detected by Angeli and Valanides (2009). When reflective discourse was promoted among a group engaged in intense Socratic questioning, the participants' critical thinking was significantly enhanced relative to a conventionally taught control group.

6.2 Relationships Between Talk Moves and Critical Thinking Operations

It was observed that evaluating moves may adversely affect students' critical thinking. Golding (2011) suggested that teachers should take a thinking–encouraging stance to scaffold critical thinking. When teachers seek predetermined answers and establish monologues through evaluating and knowledge providing moves, there are fewer opportunities for students to make authentic intellectual contributions to classroom discourse (Golding 2011). When teachers A and B staged evaluating moves frequently, student-led voices seemed interrupted.

Promotion of critical thinking is linked to a high frequency of student voices, without cutting off their remarks. This point is substantially supported by previous studies related to science education and promoting higher-order reasoning (Chin 2007; Martin and Hand 2009). Chin (2007) and Soysal and Yilmaz-Tuzun (2019) in particular concluded that, when science teachers refrained from knowledge providing and evaluating moves, students employed higher-order processes such as hypothesizing, evaluating, explaining, and deducing. This study concurs with their conclusions. Thus, when teacher C and especially teacher D abstained from evaluating moves, their students achieved higher-order critical thinking (inductive and deductive reasoning, elucidating, reflecting, creating proposals).

Communicating moves seemed essential to evoking critical thinking. Previous studies argued that critical thinking requires sustained effective communication in the classroom (Abrami et al. 2015). Golding (2011) suggested that teachers should ask “What do you mean?” questions to ascertain the underlying meaning of students’ assertions, thus ensuring healthy communication. However, he also proposed other types of talk move: why questions to guide students to give reasons (observed here as an evidencing move) or “What is the difference between...?” questions to encourage students to compare (observed here as an observe–compare–predict move). Browne and Keeley (2007) also proposed specific talk moves beyond communicating to spark critical thinking:

- What are other alternatives? (observed here as a legitimating move)
- Reasoning and elaborating questions, how might we explain more about ...? (observed here as an evidencing move)
- Evaluating questions, what are some reasons to agree with ...? What are some reasons to disagree with ...? (observed here as a legitimating move)

The implication is that communicating moves are crucial in sustaining communicatively healthy verbal interactions. This study concluded, however, that additional talk moves (legitimizing, challenging, monitoring, evidencing) are required to scaffold higher-order critical thinking. These four additional moves were found to be closely associated with higher implementation quality, and thus with promoting critical thinking.

Kuhn (1999) stressed the close relationship between taking an evaluative position by criticizing and analyzing arguments and applying higher-order critical thinking. In his view, critical thinking includes assessing and judging a given problem/opinion. In this study, these cognitive operations were more apparent when teachers staged legitimating moves, guiding students to build on shared arguments. As interpreted in the above subtopical dialogues, when legitimating moves forced students to judge and criticize others’ opinions, their arguments were revised, and students extended the limits of their own and others’ understanding. Elder and Paul (2010) and Siegel (2010) framed higher-order critical thinking as examining and criticizing propositions with a view to enhancing them. Thus, when teachers C and D enacted legitimating moves frequently, students were given the epistemic responsibility to understand, analyze, revise, and improve their classmates’ ideas.

Argument construction, analysis, and evaluation are fundamental characteristics of critical thinking (Facione 1990; Golding 2011; Siegel 1988). It is well known, however, that students are not able to analyze and evaluate arguments automatically. As shown here, these processes require intentionality and explicitness on the part of students. To spark and sustain intellectual rigor in the peer community (one of the major indicators of implementation quality), teachers had to enact legitimating moves frequently. Once teachers did so, the students were assigned

roles as coevaluators, cojudgers, and coanalyzers of proposed ideas. This discursive accountability seemed to force students to engage in reasoned discourse as a form of higher-order critical thinking, instead of striving to find an externally defined right answer. This is conceptualized as critical thinking in the strong sense (Paul 1990), and was made more concrete in the presence of frequently staged legitimating moves that seemed to boost implementation quality.

Challenging moves appeared to modify implementation quality and to evoke students' critical thinking. In Ramos' (2018) study, when teachers enacted challenging moves which valued alternative/contradictory positions, the students showed more interest in reacting to resolve conceptual conflicts. About 73% of students observed by Ramos were attentive and contributed to classroom dialogue when teachers created challenging (or argumentative) environments in which students' cognitive engagement and higher-order critical operations were maximized. King and Kitchener (1994) contended that changes in one's capacity to think critically are due to maturity and learning experiences. To create learning experiences that scaffold advanced forms of critical thinking, instructional settings must be designed to challenge students' pre-existing theories (King and Kitchener 1994) regarding, for instance, concepts in science. This suggests a close relationship between implementation quality, challenging moves, and higher-order critical thinking (King and Kitchener 1994). In teacher A's and teacher B's classrooms, challenging moves were less visible, while teacher D in particular used challenging moves at a level of at least 10%. As shown in Table 6, once teacher D performed enough discrepant talk moves, the students felt dissatisfaction with their preexisting theories on solubility. They were thus forced to transcend these theories to make sense of the concept of solubility, resulting in higher-order critical thinking (e.g., inductive reasoning).

It should be noted that joint effects of different types of talk moves can be interpreted in light of this study's outcomes. Paul (1990) stressed that better in-class implementations devoted to promoting critical thinking incorporate brainstorming activities, challenging talks (more visible in the presence of challenging moves), and accountable talks (more visible in the presence of legitimating moves) to arrive at solutions to contradictory problems. Paul's conclusions were reflected in the present study: when teachers (C, and especially D) augmented their use of challenging and legitimating moves, students appeared more stimulated to achieve higher-order critical thinking. Ennis (1993) and Halpern (2014) argued that students will judge the credibility of information sources (legitimizing moves), and defend their propositions against counter arguments (challenging moves), when they are forced to clarify, elaborate, and probe their comments (communicating moves). This was reflected in the practice of teachers C and (particularly) D, who frequently and concurrently displayed challenging, legitimating, and communicating moves. Their students seemed to attain higher-order critical thinking relative to other teachers who did not use a pragmatic combination of the above talk moves.

Rezaei et al. (2011) asserted that teacher-led questions requiring students to provide evidence may encourage them to look beneath the surface and deal with complexities. He also reported a close relationship between providing reliable evidence and thinking critically. Evidence-based reasoning should be sought by teachers, since students may lack the understanding to support their claims with examples, instances, observations, or any kind of data (Soufi and See 2019). Bell and Linn (2000) noted that students rarely base their claims on evidence. Particularly at the elementary school level, students present intuitive claims which suffer from lack of evidence. All these imply that when students are deliberately required to

support their ideas with evidence, they will provide it (Jadallah et al. 2011). In the implementations, it was observed that students were able to introduce related examples, instances, or samples to support their claims. However, such student-provided evidence was evaluated at a moderate level of critical thinking (Table 2, illustrating). To generate evidence-based reasoning in the implementations, students had to explain why they chose particular examples, or types of data, to support their claims (McNeill et al. 2006). In this study, every individualized observation, or student-proposed example/instance, was accepted as data (Table 2, illustrating). Teacher-led requests for evidencing seemed to encourage students to distinguish relevant data from irrelevant in producing a valid assertion or justified argument. In this context, supporting a claim with relevant data, or coordinating between claim and data, could be considered a higher-order critical thinking operation (Andrews 2015). It was observed that coordination was more tangible when teachers asked for supported reasoning (Table 5, line 7).

Teachers seemed to let students follow conceptual and procedural discourses as a version of metacognitive activity through monitoring moves. This teacher-directed and student-practiced metacognitive activity incorporated self-evaluations of acquisition, comprehension, and reasoning on, for instance, concepts in science (Bensley and Spero 2014; Halpern 1998; Tarricone 2011). In this study, when teachers staged monitoring moves, they directed students to think about their previous reasoning about scientific concepts (Table 7, line 1). Paul (1990) commented that one of the best ways of promoting critical thinking is to guide students in analyzing their own thinking processes.

Paul (1990) and Halpern (1998) stated that, once students begin metacognitive monitoring following talk prompts (focusing, summarizing, selecting, eliminating), this entails self-improvement and incorporation of standards for assessing thinking. Ennis (2011) and Paul (1993) contended that students' critical thinking can be best explained by self-reflective activity. Educational psychologists confirm reciprocal relationships between metacognitive monitoring, self-reflection, and critical thinking (Bensley and Spero 2014; Halpern 1998; Tarricone 2011). In this study, when teachers used monitoring moves frequently (observed especially for teacher D), the students might themselves questions such as:

- Why did the teacher direct our attention to a particular response or focal point? ("But your friend said he thought that they are the same." see also Table 7).
- Why did the teacher feature some specific responses while eliminating others? ("But now it is too confused. How does the sun warm us and other planets? Shall we talk that point?" see also Table 7).

By continuously addressing the above questions, the students monitored their own thought processes and determined if their reasoning was plausible or required refinement and improvement prior to being externalized. As observed, the teachers displayed monitoring moves to focus students on contextually more important opinions or direct them to ignore credible but contextually improper opinions. It was concluded that, when teachers increased their use of monitoring moves, students engaged in the metacognitive monitoring thought to stimulate higher-order critical thinking, as reported by previous studies (Ku and Ho 2010; Magno 2010).

It is accepted that students' observations, comparisons, and predictions, basic skills in science, are vital for in-class science inquiry activities (Kind and Osborne 2016; Kind 2013). However, when a teacher frequently asks students to make observations, comparisons, or predictions, s/he demands less cognitive effort on their part. As observed here, this may

negatively affect higher-order critical thinking. Each teacher-led talk move has a specific level of cognitive demand (Kayima and Jakobsen 2020). As Soysal (2019) examined in detail, particular talk moves may arouse specific potential cognitive demands in students according to the Bloomian taxonomy (i.e., predicting moves → understand level; communicating moves → understand level; challenging moves → evaluate level; legitimizing moves → evaluate level; monitoring moves → analyze level). On this basis, the observe–compare–predict moves observed in this study might create the following cognitive demands:

- Perception zone: Guiding students to make observations, by way of presenting relevant experiences, referring to common or different characteristics of objects, ideas, and arguments, or prompting students to make comparisons of ideas, arguments, cases, etc.
- Conception zone: Inviting students to make inferences/predictions (low level induction).

In this context, cognitive demands on students may be lower, as when teacher A relied on observe–compare–predict moves (13.78%) more than teachers C and D (1.7% and 1.3% respectively). Inviting students to make observations, comparisons and predictions was essential to the ABI approach, but these moves had to be enacted within bounds. For instance, teachers D and C used these moves, but in a limited way, accompanying them with other types of talk move (e.g., legitimating, challenging) thought to make greater cognitive demands on students (Soysal 2019), and to encourage higher-order critical thinking.

7 Conclusions

The study concluded that ABI can stimulate students to engage in higher-order critical thinking. More precisely, the ABI approach can only stimulate higher-order critical thinking if implemented with high quality. Teacher-led talk moves associated with high implementation quality appeared to evoke higher-order thinking, while those associated with lower quality had the opposite effect. Among the four teachers observed, the ability to foster critical thinking effectively was a function not only of the type of teaching but also of their intrinsic ability to implement the ABI approach. This was characterized mostly by the different types of talk moves enacted, and their proportions. When teachers dominantly enacted evaluations of student responses, or transmission of knowledge, the result was to hinder students' use of higher-order critical thinking. Frequent use of less cognitively demanding talk moves, such as observe–compare–predict, may not motivate students to display higher-order critical thinking. Timely, proper, and frequent use of challenging, legitimating, monitoring, and evidencing moves may significantly encourage students to think critically at the highest level. Communicating moves may trigger higher-order critical thinking, but this group of moves must be enacted jointly with other talk moves (challenging, legitimating, monitoring, and evidencing) found to promote students' higher-order critical thinking more effectively.

8 Educational Implications and Limitations of the Study

The ability of science teachers to keep track of relationships between quality of discourse, students' critical thinking, and types of talk moves may be questioned (Jacobs et al. 2010). Science teachers should therefore examine their own teaching practices by acting as

coresearchers (Jacobs et al. 2010). Teacher educators in particular emphasize that teachers should be able to critically analyze their own in-class practices (Chan et al., 2020). A more productive way to encourage science teachers to be reflective practitioners is to design and implement high quality, longitudinal professional development programs in which teachers can monitor, analyze, and make inferences on the quality of their in-class practices, and their effects on students' critical thinking.

The limitations of the study relate to the participants, the school science contents/concepts considered, the analytical teaching approach, and the number of teachers implementing it. Cavagnetto (2010) suggested that the effects of teaching sequences which favor argumentative discourse and in-class science inquiry (e.g., the ABI approach) on students' gains (e.g., critical thinking) should be researched, with a particular focus on earlier developmental stages (e.g., fourth/fifth graders). Soysal and Yilmaz-Tuzun (2019) reported that the effects of specific talk moves can be content-sensitive. In this study, physics and chemistry concepts were discussed. The relationships found between implementation quality, critical thinking, and talk moves, could differ for different science contents/concepts. The ABI approach can be categorized as a student-centered and process-oriented teaching strategy. The implication is that inferences derived from the current study's verbal data are more analytical and may not be transferable to other approaches to inquiry-based teaching. Finally, a study involving a larger pool of teachers might yield more generalizable findings.

Compliance with ethical standards

Conflict of Interest The author declares no conflict of interest.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, A., & Persson, T. (2015). Strategies for teaching students to think critically: a meta-analysis. *Review of Educational Research*, 85(2), 275–314.
- Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry-based approach known as the science writing heuristic to traditional science teaching practices: are there differences? *International Journal of Science Education*, 29, 1745–1765.
- Alexander, R. (2008). Culture, dialogue and learning: Notes on an emerging pedagogy. In N. Mercer & S. Hodgkinson (Eds.), *Exploring talk in schools* (pp. 91–114). London: Sage.
- Andrews, R. (2015). Critical thinking and/or argumentation in higher education. In M. Davies & R. Barnett (Eds.), *The palgrave handbook of critical thinking in higher education* (pp. 49–62). London: Palgrave Macmillan.
- Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: performance on ill-defined issues. *Learning and Instruction*, 19(4), 322–334.
- Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, 32(2), 201–211.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797–817.
- Bensley, D., & Spero, R. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity*, 12, 55–68.
- Browne, N., & Keeley, S. (2007). *Asking the right questions: a guide to critical thinking* (8th ed.). New Jersey: Pearson.
- Cairns, D., & Areepattamannil, S. (2019). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education*, 49(1), 1–23.
- Campbell, D. T., & Erdogan, I. (2005). A look at student action in the science classroom. *Science Education International*, 17(2), 101–113.

- Cavagnetto, A. R. (2010). Argument to foster scientific literacy: a review of argument interventions in K-12 science contexts. *Review of Educational Research*, 80(3), 336–371.
- Cavagnetto, A., & Hand, B. M. (2012). The importance of embedding argument within science classrooms. In M. S. Khine (Ed.), *Perspectives on scientific argumentation* (pp. 39–53). Dordrecht: Springer.
- Chan, K. K. H., Xu, L., Cooper, R., Berry, A., & van Driel, J. H. (2020). Teacher noticing in science education: do you see what I see? *Studies in Science Education*, 1–44. <https://doi.org/10.1080/03057267.2020.1755803>.
- Chin, C. (2007). Teacher questioning in science classrooms: approaches that stimulate productive thinking. *Journal of Research in Science Teaching*, 44(6), 815–843.
- Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014a). Teacher support for collective argumentation: a framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics*, 86(3), 401–429.
- Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014b). Identifying kinds of reasoning in collective argumentation. *Mathematical Thinking and Learning*, 16(3), 181–200.
- Elder, L., & Paul, R. (2010). Critical thinking: competency standards essential for the cultivation of intellectual skills, part 1. *Journal of Developmental Education*, 34(2), 38–39.
- Ennis, R. H. (1962). A concept of critical thinking: a proposed basis for research in the teaching and evaluation of critical thinking ability. *Harvard Educational Review*, 32, 81–111.
- Ennis, R. H. (1985). Critical thinking and the curriculum. *National Forum*, 65(1), 28–31.
- Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice*, 32, 179–186.
- Ennis, R. (2011). Critical thinking: reflection and perspective part II. *Inquiry: Critical thinking across the Disciplines*, 26(2), 5–19.
- Erduran, S., & Dagher, Z. R. (2014). *Reconceptualizing the nature of science for science education, scientific knowledge, practices and other family categories*. Dordrecht: Springer.
- Facione, P. (1990). *Critical thinking: a statement of expert consensus for purposes of educational assessment and instruction*. Millbrae: The California Academic Press.
- Faye, J. (2014). The nature of scientific thinking. In *On interpretation, Explanation and Understanding*. London: Palgrave Macmillan.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry based science curricula in the context of urban reform. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 45(8), 922–939.
- Glaser, E. (1941). *An experiment in the development of critical thinking*. New York: Teachers College, Columbia University.
- Golding, C. (2011). Educating for critical thinking: Thought-encouraging questions in a community of inquiry. *Higher Education Research & Development*, 30(3), 357–370.
- Grimberg, B. I., & Hand, B. (2009). Cognitive pathways: analysis of students' written texts for science understanding. *International Journal of Science Education*, 31(4), 503–521.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist*, 53(4), 449–455.
- Halpern, D. F. (2014). *Thought and knowledge: an introduction to critical thinking* (5th ed.). NY: Psychology Press.
- Hammerness, K. M., Darling-Hammond, L., Bransford, J. et al. (2005). How teachers learn and develop. In L. Darling-Hammond, & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 358–389). San Francisco, CA: Jossey-Bass.
- Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, 41(2), 169–202.
- Jadallah, M., Anderson, R. C., Nguyen-Janiel, K., Miller, B. W., Kim, I. H., & Kuo, L. J. (2011). Influence of a teacher's scaffolding moves during child-led small-group discussion. *American Educational Research Journal*, 48(1), 194–230.
- Kayima, F., & Jakobsen, A. (2020). Exploring the situational adequacy of teacher questions in science classrooms. *Research in Science Education*, 50(4), 437–467.
- Kilic, B. (2016). Investigating questioning patterns of teachers through their pedagogical progression in argument-based inquiry classrooms (Master's Thesis, Middle East Technical University, Ankara).
- Kind, P. M. (2013). Establishing assessment scales using a novel disciplinary rationale for scientific reasoning. *Journal of Research in Science Teaching*, 50(5), 530–560.
- Kind, P., & Osborne, J. (2016). Styles of scientific reasoning: a cultural rationale for science education? *Science Education*, 101, 8–31.

- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey Bass.
- Ku, K. Y., & Ho, I. T. (2010). Metacognitive strategies that enhance critical thinking. *Metacognition and Learning*, 5, 251–267.
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16–25.
- Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 371–393). Oxford: Blackwell Publishing.
- Leach, J. T., & Scott, P. H. (2002). Designing and evaluating science teaching sequences: an approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science Education*, 38, 115–142.
- Leach, J. T., & Scott, P. H. (2003). Individual and sociocultural views of learning in science education. *Science & Education*, 12(1), 91–113.
- Lefstein, A., Snell, J., & Israeli, M. (2015). From moves to sequences: expanding the unit of analysis in the study of classroom discourse. *British Educational Research Journal*, 41(5), 866–885.
- Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning*, 5, 137–156.
- Mameli, C., & Molinari, L. (2014). Seeking educational quality in the unfolding of classroom discourse: a focus on microtransitions. *Language and Education*, 28(2), 103–119.
- Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, 39(1), 17–38.
- McDonald, C. V., & McRobbie, C. J. (2012). Utilising argumentation to teach nature of science. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 969–986). Dordrecht: Springer.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, 15(2), 153–191.
- McPeck, J. E. (1981). *Critical thinking and education*. New York: St. Martin's Press.
- McPeck, J. E. (1990a). Critical thinking and subject specificity: a reply to Ennis. *Educational Researcher*, 19(4), 10–12.
- McPeck, J. E. (1990b). *Teaching critical thinking: dialogue and dialectic*. New York: Routledge.
- Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, 80(1), 1–14.
- Molinari, L., & Mameli, C. (2013). Process quality of classroom discourse: pupil participation and learning opportunities. *International Journal of Educational Research*, 62, 249–258.
- Molinari, L., Mameli, C., & Gnisci, A. (2013). A sequential analysis of classroom discourse in Italian primary schools: the many faces of the IRF pattern. *British Journal of Educational Psychology*, 83(3), 414–430.
- Moore, T. J. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development*, 30(3), 261–274.
- Norton-Meier, L., Hand, B., Hockenberry, L., & Wise, K. (2008). *Questions, claims, and evidence: the important place of argument in children's science writing*. Portsmouth: Heinemann.
- Paul, R. W. (1990). *Critical thinking: what every person needs to survive in a rapidly changing world*. Santa Rosa: Foundation for Critical Thinking.
- Paul, R. W. (1993). *Critical thinking: What every person needs to survive in a rapidly changing world*. Santa Rosa, CA: Foundation for Critical Thinking.
- Paul, R., & Elder, L. (2008). Critical thinking: strategies for improving student learning, part II. *Journal of Developmental Education*, 32(2), 34–35.
- Phillips, F., & Mackintosh, B. (2011). Wiki art gallery Inc.: a case for critical thinking. *Issues in Accounting Education*, 26(3), 593–608.
- Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed Teaching Observation Protocol (RTOP)*. ACEPT IN-003. Arizona: Arizona collaborative for excellence in the preparation of teachers, Arizona State University.
- Pimentel, D. S., & McNeill, K. L. (2013). Conducting talk in science classrooms: investigating instructional moves and teachers' beliefs. *Science Education*, 97(3), 367–394.
- Poock, J. A., Burke, K. A., Greenbowe, T. J., & Hand, B. M. (2007). Using the science writing heuristic to improve students' academic performance. *Journal of Chemical Education*, 84, 1371–1379.
- Ramos, J. J. R. (2018). Critical thinking skills among senior high school students and its effect in their academic performance. *International Journal of Social Sciences & Humanities*, 3(2), 61–73.
- Rezaei, S., Derakhshan, A., & Bagherkazemi, M. (2011). Critical thinking in language education. *Journal of Language Teaching and Research*, 2(4), 769–777.

- 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.
- Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. New York: Routledge.
- Siegel, H. (2010). Critical thinking. In E. Baker, B. McGaw, & P. Peterson (Eds.), *International encyclopedia of education* (Vol. 6) (pp. 141–145). Amsterdam: Elsevier Academic.
- Soufi, N. E., & See, B. H. (2019). Does explicit teaching of critical thinking improve critical thinking skills of English language learners in higher education? A critical review of causal evidence. *Studies in Educational Evaluation*, 60, 140–162.
- Soysal, Y. (2019). Investigating discursive functions and potential cognitive demands of teacher questioning in the science classroom. *Learning: Research and Practice*, 1–28. <https://doi.org/10.1080/23735082.2019.1575458>.
- Soysal, Y., & Yilmaz-Tuzun, O. (2019). Relationships between teacher discursive moves and middle school students' cognitive contributions to science concepts. *Research in Science Education*. <https://doi.org/10.1007/s11165-019-09881-1>.
- Stark, E. (2012). Enhancing and assessing critical thinking in a psychological research methods course. *Teaching of Psychology*, 39(2), 107–112.
- Sternberg, R. J., & Williams, W. M. (2002). *Educational psychology*. Boston: Allyn and Bacon.
- Tarricone, P. (2011). *The taxonomy of metacognition*. New York: Psychology Press.
- Tirunch, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: a systematic review of intervention studies. *Higher Education Studies*, 4(1), 1–17.
- Tirunch, D. T., De Cock, M., Weldeslassie, A., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663–682.
- Tirunch, D. T., De Cock, M., & Elen, J. (2018). Designing learning environments for critical thinking: examining effective instructional approaches. *International Journal of Science and Mathematics Education*, 16, 1065–1089.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toy, B. Y., & Ok, A. (2012). Incorporating critical thinking in the pedagogical content of a teacher education programme: does it make a difference? *European Journal of Teacher Education*, 35(1), 39–56.
- Tytler, R., & Aranda, G. (2015). Expert teachers' discursive moves in science classroom interactive talk. *International Journal of Science and Mathematics Education*, 13(2), 425–446.
- van Booven, D. (2015). Revisiting the authoritative–dialogic tension in inquiry-based elementary science teacher questioning. *International Journal of Science Education*, 37(8), 1182–1201.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Science & Education is a copyright of Springer, 2021. All Rights Reserved.