Fall 2015

An analysis of teacher question types in inquiry-based classroom and traditional classroom settings

Sungho Kim
University of Iowa

AN ANALYSIS OF TEACHER QUESTION TYPES IN INQUIRY-BASED CLASSROOM AND TRADITIONAL CLASSROOM SETTINGS

by

Sungho Kim

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy
degree in Teaching and Learning (Science Education) in the
Graduate College of
The University of Iowa

December 2015

Thesis Supervisor: Professor Brian Hand

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Sungho Kim

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy degree
in Teaching and Learning (Science Education) at the December 2015 graduation.

Thesis Committee: ___________________________________
Brian Hand, Thesis Supervisor

____
Soonhye Park

____
Mark McDermott

____
Earnest Pascarella

____
Ingrid Peate

To my parents, wife, and son, I dedicate this dissertation for their unconditional love and endless support. I love you dearly

ACKNOWLEDGEMENTS

I would like to express my appreciation to my loving wife, Youngnam, my son, Jooyoung, my parents, Dr. Brian Hand and Mr. Jay Harding for the preparation of my dissertation.

ABSTRACT

This study examined the differences and patterns for three categories between an argument-based inquiry group and a traditional group over the period of the SWH (Science Writing Heuristic) project: (1) teacher talk time, (2) structure of questions (question types), and (3) student responses. The participating teachers were chosen randomly by a convenient sampling method because the data were collected previously from the SWH project. Each group had thirty teachers. A total of sixty teachers participated in the study. Student responses were part of the study to evaluate the effect of open-ended question types but students were not direct participants in the study. Each teacher was asked to send a recorded video clip of their class at the end of each semester (spring and fall) over two years. Each teacher sent four video clips for the project. A total of two hundred forty video clips were analyzed to gather the information regarding the three categories. The first category was teacher talk time. It was measured in seconds only when teachers interacted with students with the topic. The second category was the structure of questions (question types). It consisted of two question types (open-ended and close-ended). Under the open-ended question category, there were three sub-question types: (1) asking for explanation (AE), (2) asking for self-evaluation of reasoning (AF), and (3) asking for self-evaluation of others' reasoning (AFO). Under the close-ended question category, there were two sub-question types: (1) asking for factual information (AI) and (2) asking for confirmation (AC). Each sub- question type was counted numerically. The last category was student responses. Student responses consisted of higher-order thinking and lower-order thinking. Under the higher-order thinking category, there were three sub-types: (1) explanation responses (E), (2) self-evaluation of reasoning responses (SE), and (3) self-evaluation of others' reasoning responses (SEO). Under the lower-order thinking category, there was one sub-type: simple responses (S).

Each sub type was counted numerically. Based on the descriptive results (the length of teacher talk time in seconds, the number of question types, and the number of student responses), repeated measures ANOVA was conducted to find any differences and patterns for teacher talk time, structure of questions and student responses between the treatment and control groups over the period of the project and across time (four different time points). The results showed that there were clear differences for teacher talk time, the structure of questions, and student responses between the treatment and control groups over the period of the project and at each time point. The treatment group teachers talked less and used more open-ended questions than the control group teachers. The treatment group students displayed more higher-order thinking responses than the control group students.

PUBLIC ABSTRACT

This study examined the differences and patterns for teacher talk time, question types, and student responses between argument-based inquiry classes and traditional classes. In each group, thirty teachers participated in the study. A total of sixty teachers' classes participated in the study over two years. Student responses were part of the study to evaluate the effect of open-ended question types but students were not direct participants in the study.

The analyses were conducted using statistical tests (repeated measures ANOVA and Cohen's d). The results showed that there were clear differences between the argument-based inquiry classes and traditional classes. The argument-based inquiry class teachers talked less, asked open-ended questions more frequently than the traditional class teachers over the period of the study. The argument-based inquiry class students displayed higher-order thinking responses more frequently than the traditional class students over the period of the study.

Based on the statistical results, specific patterns emerged. Teachers talked less, used more open-ended questions and students responded with a higher frequency of higher-order thinking in the argument-based inquiry classes. In order to find the degree of the differences between the two groups, effect sizes were calculated. The degree of the differences between the two groups was greater than medium.

The study suggested that teachers should be encouraged to talk less and use open-ended questions to elicit higher-order thinking student responses when applying an argument-based inquiry approach in elementary school science classes for student learning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

CHAPTER 1

INTRODUCTION

Inquiry has been a major theme in science education reform [National Research Council (NRC), 2012]. Although there has been considerable research on inquiry many teachers have difficulties implementing inquiry in the classroom especially at the elementary level. Teaching science in an inquiry-based class is a more daunting task for elementary school teachers than for secondary teachers because elementary teachers in general lack confidence in science content. Davis and Smithey (2009) explained this phenomenon by saying "elementary teachers face further challenges, since at the elementary level, teachers are responsible for life science, physical science, and earth science." According to Fulp (2002) and Weiss et al (2001), elementary teachers showed the tendency to feel less comfortable when teaching science and to put less effort into preparing and teaching science than other subjects. This may show that elementary teachers do not possess enough content knowledge to teach science based on inquiry with confidence (TIMSS, 2003).

This leads to challenges for teaching science with inquiry. Inquiry-based classrooms are organized so that students have the opportunity to work together on open-ended problems and talk actively to solve problems (Kelly & Brown, 2003; Roychoudhury & Roth, 1996; Scott, 2005). In order to encourage teachers to apply inquiry in the classroom, the role of language in science education has drawn attention, particularly in the area of discourse (Candela, 1999; Lemke, 1990; Mortimer & Scott, 2003; Ogborn, Kress, Martins & Mcgillicuddy, 1996; Aguiar, Mortimer & Scott, 2009; Scott, 1998; Sutton, 1992). The broad definition of the term 'discourse' is understood as the usage of language within social contexts (Gee, 2001). However, in the context of science education research, the concept of discourse contains more complicated components to clarify its meaning. Gee (1996) explained discourse as an interaction between ‘‘words, acts, values, beliefs, attitudes, and social identities’’ (p. 126) within groups of individuals who contribute mutually to sense-making and development of meaning through argumentation. According to the definition of discourse, it is not sufficient to define discourse as equivalent to classroom talk. Discourse is a complicated interaction between the participants. It is a form of demonstrating the teacher's and students' distinct perspectives via oral communications.

Among many components in language, questioning occurs frequently between teacher and students, and between students. Questions are the most prevalent teaching practice. Teachers ask questions during class to help students learn the topic. Graesser & Person (1994) found that the frequency of teacher questions is contingent on the type of activity, and ranges between 30 and 120 questions per hour. Additionally, Roth (1996) summarized the pattern of questioning as follows:

> A survey of several studies reveals that in the science classroom, the rate of teacher questions is related to activity type and teacher knowledge (Carlsen, 1991). The questioning rate is highest during lectures (82 questions per hour) and lowest during routine seat work (3 questions per hour), and the rate of questioning is negatively correlated with teacher subject-matter knowledge.

These studies showed that questioning occurs in the classroom regardless of the type of activity and class. Interestingly, the frequency of questioning is not related to teacher subject-matter knowledge. This means regardless of how much content knowledge the teacher possesses, they use questions in the classroom as a strategy for facilitating student learning. Questions play a central and fundamental role in learning, and questions generated by teachers are indicators of quality teaching (Carlsen, 1993; Roth, 1996; Smith, Blakeslee, & Anderson, 1993). It is important to understand what components questions should include in order for students to have meaningful learning. Roth (1996) elucidated those components of questions based on King's (1994) work, in which good questions are defined as facilitating learners' thoughts centered on their experience and subsequently leading to creative thinking, as follows;

> To bring about conceptual change in science, good teachers use questions to elicit student explanations, elaborations of previous answers and ideas, and predictions that contradict students' intuitive ideas about natural phenomena (p. 710)

However, the type of questions most teachers ask in the classroom is to demand explicit, factual information rather than student reasoning. In this sense, the teachers do not take advantage of the potential of good questioning for leading students to meaningful learning (Graesser & Person, 1994).

Even though, a number of research studies about questioning have been conducted, the essence of questioning has been difficult to assess since questioning contains complexities (Chin, 2007; Roth, 1996). Carlsen (1991) shed light on the investigation of teacher questioning by designing a framework that focused on three essential features: (1) the context of questions, (2) the content of questions, and (3) the responses and reactions to questions. First, the context of questions can be characterized by two factors: the situation in which the person asks questions and the way that people ask questions. Researchers are required to pay attention to how the questioner constructs the historical, physical, and social planes of the situation; and both the verbal and nonverbal practice the questioner has been and will be using (Ochs, 1979). Second, a small number of research studies have focused on the relation between questions and the depth of content knowledge that students achieve. This type of study is called process-product.

Research on process-product (e.g., Dantonio & Paradise, 1988; Mill, Rice, Berliner & Rousseau, 1980; Redfield & Rousseau, 1981; Winne, 1979**)** has stratified questions based on cognitive levels but the empirical results were questionable because researchers used different coding systems. Additionally, literature reviews and meta-analyses of research using this method have not shown agreement about the results. They have shown different results and interpretations about their studies (Chin, 2007; Redfield & Rousseau, 1981; Samson, Strykowski, Weinstein, & Walberg, 1987; Winne, 1979). In contrast, research that considered sociolinguistic perspectives has specific strengths, such as providing a way to describe the context of handling content knowledge.

However, it also has significant limitations. For example, this type of research requires (a) enormous time commitment, (b) depth and breadth of subject-matter knowledge, and (c) complicated analyses. Due to these reasons, this type of research is not prevalent in science education (e.g., Carlsen, 1988; Lemke, 1990; Roth, 1996). However, there was one important finding from these studies: teachers tended to take control of classroom discourse by framing, validating, and sequencing the content of knowledge. This instructional approach was identified as the Initiation-Response-Evaluation [IRE] sequence that is still commonly observed in the classroom: The teacher initiates (I) interactions with questions, a student responds to the teacher's questions (R), and the teacher evaluates the responses immediately (E). The teacher's questions that elicit lower-order thinking usually occur at the recall session. This process is also known as the IRF (initiation, response, and follow-up) method (Sinclair & Coulthard, 1975) because the last part does not necessarily need to be an evaluation. In other words, these studies have primarily focused on the difference between close-ended, expected answer questions intrinsic in the initiation-response-evaluation [IRE] or recitation pattern (Cazden, 1988; Mehan, 1979; Sinclair & Coulthard, 1975) and open-ended questions intended to evoke and scaffold student reasoning (Smith, Blakeslee, & Anderson, 1993). Later, Mortimer and Scott (2003) expanded the IRE sequence by adding more of an additional couplet of response and feedback into a more expounded initiation-response-feedback-response-feedback [IRFRF] sequence structure in which expatiated feedback from the teacher elicits a more detailed or thoughtful response from the student.

Third**,** the main measuring factor of process-product research is “wait time” for students’ responses and reactions to questions. However, sociolinguistic approaches for questioning pay more attention to different factors than wait time, such as social status, participating patterns, and quality of responses (Carlsen, 1991)**.** Previous research on teacher questioning indicated that teacher questions usually elicit terse and factual statements on the part of students. It explains how teacher questioning might discourage students learning. For example, (1) teacher talk time (i.e., lecturing) is prevalent throughout class rather than student talk time, (2) types of questions are directed at eliciting factual information, and (3) students do not have the autonomy to discuss any topic in class (Carlsen, 1988). Alternatively, student talk time increases when teachers ask personal questions and show sincere interest, do not assess students’ responses and students take control of the class discussions (Roth, 1996).

Purpose

Teacher questioning has been considered one of the most effective methods to facilitate meaningful student learning. There have been many efforts to understand the essence of teacher questioning but it is a daunting task (Chin, 2007; Roth, 1996). The goals of this study are to contribute to understanding of how teacher questioning associated with teacher talk time affects student learning in terms of student responses and how teachers adopt the usage of questions in class to promote students’ understanding of a concept. This study took place in the context of a field trial in which the researcher manipulated the participants by assigning them into treatment and control groups based on schools. During the field trial, the researcher investigated and analyzed how teacher questioning combined with the length of teacher talk time was utilized in relation to student responses in two different classroom settings (i.e., argument-based inquiry classes and traditional classes). There is a concern that the potential of questioning is not being fully utilized as intended and advocated in the classroom (Downing & Gifford, 1996), particularly for questioning that expedites and scaffolds student reasoning for the inquiry learning process (Martin & Hand, 2009; Weiss & Pasley, 2004; Carlsen, 1997). Different types of teaching (e.g., inquiry-based class or traditional class) induce different discourse patterns (Van Zee, Iwasyk, Kurose, Simpson, & Wild, 2001). In a traditional classroom, the teacher explains facts and procedures, and students sit and listen without interacting with the teacher. During lectures, the teacher may ask or answer a question but student questions are minimal (Dillon, 1988a). Instruction is mainly didactic through ''teaching by telling'' and the primary purpose is to disseminate information in an efficient way within a limited time. In order for teacher questioning to be more effective, Chin (2007) and Mortimer and Scott (2003) explained that the role of teachers in the science classroom is to put students within the social language of science. The teacher must introduce scientific ideas to students by providing a social context and encourage students to apply the ideas to other situations. In doing so, the students are eventually able to internalize and make sense of those ideas. Additionally, Chin (2007) elaborated on this approach of using questions as follows:

> During a guided discussion, the teacher asks conceptual questions to elicit students' ideas and facilitate productive thinking, invites and welcomes students' responses and questions, provides on-going assessment by commenting on students' responses, and encourages multiple responses. All these are done with the aim of helping students construct knowledge in the spirit of inquiry and constructivism (e.g., Roth, 1996; Settlage, 1995; van Zee & Minstrell, 1997a,b).

This study was intended to clarify the role of teacher questioning in science classes and the effective usage of questioning for student learning in the inquiry environment. The participants of this study were teachers who participated in a Scientific Writing Heuristic [SWH] project for three years. The teachers were divided into one experimental and one control group. Teacher questioning from each group was analyzed to see if there was a difference between teachers in argument-based inquiry classes and traditional classes in term of teacher talk time, types of questions, types of responses to questions and how the teachers used a sequence of questions to promote student learning.

Research questions

This study was designed to investigate how teacher questioning plays a role in enhancing student learning. Among many components in language, questioning is the most prevalent and natural component that has been practiced from either the part of a teacher or a student. Questioning is perceived as a strong tool to elicit student reasoning when learning a new concept (Chin, 2007). Additionally, teachers tend to dominate class time by asking many questions without providing enough wait time rather than allowing their students to talk (Carsen, 1991). As mentioned above, many researchers have focused on teacher questioning for this reason. There are four main questions raised for the study.

- Is there a difference in the amount of teacher talk time between teachers in argument-based inquiry classes and teachers in traditional classes?
- Are there differences in question types and student responses between argument-based inquiry classes and traditional classes?
- What are the differences in teacher talk time, teacher questioning and student responses between argument-based inquiry classes and traditional classes?
- What is the relation between teacher talk time, teacher questioning and student responses within each group?

These questions guide the direction of the study about the patterns of teacher talk time, types of questions (open-ended or close-ended questions), and student responses (higher order thinking or lower order thinking responses) between teachers in argument-based inquiry classes and teachers in traditional classes and relations of teacher talk time, types of questions, and students responses within each group. In recent studies, open-ended questioning or scientifically oriented questioning has emerged as a powerful psychological instrument in inquiry learning that helps learners to critique one's or other's ideas (Oliveira, 2010; NRC, 2000, 2012). It is important to investigate the relation between types of questions and types of responses since the relation might be an indicator of what type of questions can facilitate student responses that are advocated by NRC (1996, 2000). During the inquiry process, teacher questioning is designed to encourage students to explain and argue their answers with logic (e.g., higher order thinking), not to evaluate whether the responses are right or wrong in the form of being asked to provide factual information (e.g., lower order thinking). Finishing the question sequence in an assessment remark is not desirable as students should be required to self-evaluate their responses based on evidence and justify their claims by providing evidence (Morge 2005). To put it simply, the teacher should relinquish the assessment role back to the students. It leads the teacher to creating an atmosphere that promotes justification, conjecture, and the development of knowledge through critique (Chin, 2007; Smart & Marshall, 2013). Question sequence identifies how teachers administer questions to promote student learning. Teachers are encouraged to ask questions according to the status of student understanding of the topic (Chin, 2007). Lustick (2010) also argued that the purpose of employing specific types of questions during inquiry-based classes is to generate questions that focus on being supportive about student understanding of a concept while students take part in the process of scientific inquiry.

The rest of the chapters will be presented in terms of literature review (chapter 2), method (chapter 3), results (chapter 4), and discussion and implications (chapter 5). The literature review chapter will discuss the theoretical foundations of teacher questioning and student responses for the analytical framework of the study based on previous research. The method chapter will discuss the process of analyzing and interpreting data using statistical tests (e.g., repeated measures ANOVA and Cohen's d). The results chapter will provide the interpretation of results based on statistical tests. The discussion and implications chapter will provide the discussion of the results presented in chapter 4, answers to the research questions, the findings that emerged through the interpretation of the results. This chapter concludes with limitations, implications and future directions for the study.

CHAPTER 2

LITERATURE REVIEW

This chapter will address six main themes regarding teacher questioning. The first theme is the importance of teacher questioning. The second theme is the development of teacher questioning research. The third theme is the methods teachers use to navigate obstacles to appropriate question use in argument-based inquiry classes. The fourth theme is the effect of teacher questions on student reasoning during the learning process. The fifth theme is student responses in regard to teacher question types. The fifth theme is new directions for researching teacher questioning. The last theme is emerging analytical framework for the study.

The importance of teacher questioning

Teacher questions have long been regarded as an important factor in teaching and student learning (Chin & Osborne, 2008; Graesser & Person, 1994; Lee & Kinzie, 2012; Roth, 1996). Because of this reason, there have been a number of research studies regarding many aspects of teacher questioning. The section on teacher questions is divided into three parts: the context of questioning for argumentation, definition of high-quality questions and theoretical foundations of questioning.

Context of questioning for argumentation

Argumentation has been advocated by many researchers (e.g., Erduran & Osborne, 2005; Kim & Hand, 2015; Pontecorvo, 1987; Schwarz et al., 2003). Argumentation is a form of discourse that helps students understand new concepts critically rather than accepting them as they are without any critique (e.g., memorizing concepts without understanding). Through this process, students are able to develop a concrete understanding of the new concepts because they go through confirmation stages through arguments (Zohar & Nemet, 2002; Jim´enez-Aleixandre

& Erduran, 2008). This explanation aligns with the concept of inquiry. Teacher questions should encourage arguments from students in order for students to engage in inquiry. The ability for students to study valid knowledge on their own (NGSS, 2013) has been a main theme in science education.

Definition of high-quality questions

High-quality questions are grounded in students' experiences and lead to creative thinking. High-quality questions are designed to promote students' learning and develop their thinking process (Lee & Kinzie, 2012). A number of studies regarding teacher questioning were conducted using analytical frameworks that include the definition of high-quality questions based on Bloom's taxonomy to investigate the effects of high-quality questions on student learning. According to Massey et al. (2008), high-quality questions include involving students in extended cognitively-challenging discourse. Many researchers asserted that high-quality/open-ended questions should be used in inquiry-based classes to assist students in cultivating higher-order language skills such as predicting, analyzing, and inferring (Wasik et al., 2006). Other researchers, such as Harlen and Qualter (2004), proposed that teachers are required to ask open-ended questions in inquiry-based classes because they elicit student reasoning (e.g., analysis and reflection) while they are learning new concepts.

Moreover, open-ended questions have been advocated as stepping stones for enriching vocabulary and creating an atmosphere where students feel comfortable sharing their ideas with their peers and teachers in class compared to close-ended questions (de Rivera et al., 2005, Peterson et al., 1999, and Whitehurst et al. 1994). Chin (2007) also argued that questions are a crucial component of classroom discourse. This means that the questions asked by teachers have "potential as a psychological tool" in helping construct students' knowledge. Classroom discourse may be dichotomous: authoritative and dialogic (Scott, 1998). In authoritative discourse, the teacher focuses on disseminating information to the student. The student's answers are usually "yes or no" or short answers rather than responses that require an analytic thinking process. In contrast, dialogic discourse requires the teacher to encourage students to engage in activities using their own reasoning processes. In other words, students are able to generate, explore, investigate, and debate their own ideas. This indicates that "when students engage in dialogic discourse that fosters more generative thinking, good habits of mind such as questioning and relating ideas are rehearsed on the social plane." (Chin, 2007, p. 816). Scott (1998) also maintained that this might lead to laying the foundation of "the basis of active, analytic, individual thought."

Open-ended questions can investigate the abstract concepts of an argument, lead to unexpected contexts or other possible arguments, or set the stage for the validity of an argument (Walton & Godden, 2005). They are able to serve as stepping stones for rebuttals of one's own or others' opinions. Open-ended questions are able to be used either in a lower cognitive sense (remembering and/or understanding in terms of exploring further into the argument) or in a higher cognitive sense (evaluation in terms of a challenge or objection to the argument). From this perspective, they play a role as heuristic devices to promote dialectical reasoning. Particularly when the topic at issue that needs to be settled by argumentation requires fair views of consideration. Overall, high-quality (open-ended) questions are defined as questions that elicit student reasoning in terms of requiring students to analyze, predict, evaluate and create ideas when learning new concepts.

Theoretical foundation

Social development theory by Vygotsky (1978) has been used to explain the theoretical foundation for teacher questioning. Bransford et al. (1999) interpreted Vygotsky's theory as a means of understanding the effect of teacher questioning on student learning because students do not learn in isolation from social contexts (e.g., dialogues with others). For example, interactions with more experienced, knowledgeable or adept people (e.g., parents, teachers, peers, etc.) lead students to construct a better understanding of the concept. This idea is explained by the Zone of Proximal Development (ZPD). It elucidates individual development as the distance between students' abilities to undergo a task under the guidance of an adult and/or with the collaboration of a peer and the students' abilities to solve the problem on their own. Bruner (1966) termed a teacher's help within an individual student's ZPD as ''scaffolding''. For example, when a student encounters a daunting task that they are not able to resolve on their own, a teacher would be able to effectively scaffold the student's problem-solving by motivating them to use alternative strategies within their ZPD range such as showing pictures for clues instead of telling the student the correct answer immediately (Bodrova & Leong, 1998).

Teacher questioning plays a role in helping students move to the next cognitive level. Particularly, open-ended questions lead students to realize what they know and what they do not know because open-ended questions require divergent answers (i.e., multiple answers) compared to closed-ended questions which require convergent answers (i.e., one correct answer). This means open-ended questions promote student reasoning and do not pressure students to respond with a single right answer. With open-ended questions, students are able to learn knowledge through trial and error and derive knowledge using argumentation components (e.g., challenging, supporting, or rejecting) and Bloom's revised taxonomy higher position components (e.g., creating, analyzing, or applying). Through this process, students are able to realize what they know and what they do not know for themselves. So, they are able to acquire knowledge by correcting their misunderstanding on their own. Open-ended questions assist students in realizing how to learn knowledge on their own because they provide students opportunities to reason ideas through argumentation. In contrast, close-ended questions do not lead students to move to the next cognitive level because they emphasize memorizing or reiterating knowledge without utilizing the reasoning process. Student learning takes place when students move up to the next cognitive level with the help of teacher open-ended questioning. Overall, the Zone of Proximal Development [ZPD] illustrates how teacher questioning should be structured for student learning.

The other theoretical foundation that is relevant to this study is contained within the information processing model. The information processing model (Atkinson & Shiffrin, 1968) explains the process of how students respond to a teacher's questions. Responding to a question requires multiple steps; (1) listening to a question, (2) understanding the intent of the question, and (3) engendering a response (Gall, 1984).

The first step is for the student to respond to the question that is asked. In this process, the filter theory (Broadbent 1958) can be applied. This theory assumes that a filter or bottleneck in processing sensory information appears when people are listening to a question and proposes that people choose information to process based on its physical properties such as the tone of the person who delivers a question. In other words, people decide to filter which of the incoming auditory information to process. The remaining unattended auditory information is deleted from the listener's attention and, eventually, not processed.

The second step in the cognitive process is that people comprehend the question. Students interpret the intent of a question by recognizing individual words within the context of the question, recalling lexical information about those words, and converting the acquired information into a conceptual representation of their overarching and synthesized meaning (Anderson, 2004). Once students comprehend a question, they produce a response by evaluating related knowledge stored in their memories or available in class resources and materials, and then selecting relevant information to utilize in their responses (Gall, 1984).

This process indicates how teacher questioning should be utilized to help students move up to the next cognitive level. In sum, teacher questioning should consider the process of student learning as well as the process of students' understanding and responding to teacher questions. Based on these theories, Bloom's revised taxonomy (Anderson et al, 2001) along with previous research about question types and student response types were used to create an analytical framework for the study.

The development of teacher questioning research

Types of question

Question types have been divided into two to six types by scholars. Initially, researchers focused on four question types. Blosser (1973) analyzed teacher questioning into four types: (1) open-ended question (e.g., "If you were to design a science display for the school bulletin board, what would you include in the display and why?" or "What should be included in a project to improve the school environment?"), (2) closed-ended question (e.g., "What is the chemical formula for water?" or "What happened when you switched from low to higher power magnification?"), (3) rhetorical question (e.g., "The green coloring matter in plants is called chlorophyll, right?" or "Yesterday we said there are three major groups of rocks, right?"), and (4) managerial question (e.g., “Does everyone have the necessary equipment?” or “Will you turn to page 15?”). Wajnryb (1992) divided teacher questioning into six question types: (1) yes/no questions: teacher knows the answer (e.g., “Is this right?”), (2) short answer/retrieval-style questions: reiterating (e.g., “Could you tell me what this is?”), (3) open-ended questions: being divergent (e.g., “What needs to be done to write a report?”), (4) display questions: using representations (e.g., “What is it?” using flash cards), (5) referential questions: teacher does not know the answer (e.g., “Did you read a book to find information?”), and (6) non-referential questions: asking for an opinion or judgment (e.g., What do you think about this topic?”).

However, as research advanced, scholars started to focus on two question types (open-ended and close-ended). This trend was supported by Chin (2007), King (1994), NRC’s definition of inquiry (2012) and many researchers. King (1994) and Wragg and Brown (2001) also defined desirable question types in their research in terms of open-ended questions, which should help elicit student reasoning when students learn new concepts. Chin (2007) analyzed previous research regarding question types (e.g., Baird & Northfield, 1992; Lemke, 1990; Mehan, 1979; Mortimer & Scott, 2003; van Zee & Minstrell, 1997b) and summarized two question types which are used in traditional and inquiry-based classes.

Table 2. 1. Comparison of teacher questioning in traditional and constructivist teaching

	Traditional	Constructivist/Inquiry
Purpose of questioning	Evaluate what students know	Elicit what students think, encourage them to elaborate on their thinking, and help them construct conceptual knowledge
Structure of questioning sequence	IRE (teacher-student-teacher)	IRFRF chain
Adjustments to teachers' agenda	Move through a series of questions in accordance with planned agenda	Reflective toss (student-teacher-student) Adjust questioning to accommodate students' contributions and respond to students' thinking
Nature of questions and responses	Recall, lower order, closed with predetermined short answer	Open, engage students in taking more responsibility for thinking (higher-order thinking); responses are longer, calling for one- or two-sentence answers
Teacher's response	Praise correct answers; correct wrong answers; treat students' challenges to their questions as threat	Delay judgment: accept and acknowledge student contributions in a neutral rather than evaluative manner

Issues in utilizing teacher questions in argument-based inquiry classes

Researchers have found that most teachers do not use open-ended questions frequently in their teaching practice (Galton et al. 1999; Walsh and Sattes 2005). In one research study, Wragg and Brown (2001) observed more than a thousand questions asked by teachers in primary grades and categorized their questions as: managerial (questions which require the running of the lesson), information/data (questions which ask for recalling information), and higher order (questions which asked students to go beyond the simple recollection of facts, encouraging them to analyze, generalize, or infer information). Their results found that most questions focused on classroom management and factual recall. Only a few questions encouraged students' higher order thinking. This insufficiency of high-quality (open-ended) questions may not promote a creative environment which makes the most of students' cognitive and linguistic abilities.

In another research study, the IRE sequence (Mehan, 1979) or “triadic dialogue” (Lemke, 1990) has been investigated in other fields with the same name or a different name. The IRE sequence is still frequently used in the science classroom even though many studies have expressed concern that this type of question sequence does not elicit student reasoning (Chin, 2007; Roth, 1996) and may even discourage student learning by not allowing students enough time to think about concepts on their own due to an immediate evaluation from the teachers. McNeill and Pimentel (2010), and Lemke (1990) identified that teachers used close-ended questions that require fixed answers frequently in the classroom. This line of research implied that teachers had difficulty changing their practice into argument-based inquiry practice.

However, the research conducted by Martin and Hand (2009) shed some light on this issue. They found that one teacher changed their discourse pattern over two years into argument-based inquiry discourse patterns in terms of posing questions from close-ended to open-ended questions. At the beginning of the project, the teacher used more closed-ended questions (e.g., recall of information). Later, the teacher used open-ended questions (e.g., allowing multiple answers) more. That practice induced a higher percentage of students using reasoning (e.g., student responses with evidence for their claims and rebuttals for ideas).

On the other hand, in classes which put an emphasis on dialogue (Lemke, 1990) and the change of concepts with constructivist-based instruction approaches (Smith, Blakeslee, & Anderson, 1993); teacher questioning emphasizes different aspects compared to traditional classes. In this type of class, the teacher tends to give more freedom to their students while studying science. For example, students are encouraged to generate, explore, predict, and debate ideas and questions, even if their observations and conclusions are different from the teacher's.

This type of questioning is open-ended and requires more higher-order thinking (Baird & Northfield, 1992). Harlen and Qualter (2004) asserted that teachers' questions play a pivotal role in every step of students' scientific investigations. The overall function of questions (1) stimulates students' explorations of scientific phenomena and development of process skills, (2) scaffolds students' scientific investigations, and (3) encourages them to relate the evidence collected to the concept. Van Zee and Minstrell (1997b) invented the term; the ''reflective toss,'' which consists of a three-step structure: student statements, teacher questions, and additional student statements. Chin (2007) interpreted that by saying "the toss metaphor suggests a teacher ''catching'' the meaning of the student's prior utterance and ''throwing'' responsibility for thinking back to the student and all those present in class."

Teacher questions have been studied by many researchers because teacher questions are regarded as an important component of student learning. Early research on teacher questioning focused on types of questions and frequency of questions that was used in class rather than the way teacher questioning effects student learning. In other words, researchers focused on the structure of questioning itself rather than how it interacted with student responses and other factors (e.g., teacher talk and student talk time). The trend for researching teacher questioning has recently shifted from the identification of structure of questions to the interaction between teacher questions and student learning in terms of evaluating student response types. This current trend provides practical instructions to teachers who struggle to implement argument-based inquiry into their classes.

The methods teachers navigate obstacles to using questions appropriately in argument-based inquiry classes

Lee and Kinzie (2012) asserted that diverse classroom contexts are able to limit a teacher's use of questions in regard to question types (open-ended and close-ended questions) and their frequency (e.g., teacher talk time). In order to resolve obstacles teachers encounter while using questions in argument-based inquiry classes, numerous studies suggested that the role of the teacher needs to be defined. Previously, Dean (1986) asserted that "one must keep in mind that the nature of the question has a remarkable impact on the progression of thought in the class" (p.185).

The role of teacher

Scott (1998) argued that authoritative discourse has its own role in the classroom. When teachers use these two types of discourse (authoritative and dialogic) in the classroom in an alternated and balanced fashion (i.e., a "rhythm of discourse"), learning will be more effective. Crawford (2000) also emphasized that teachers should not pressure students to respond with specified answers to their questions. This may lead to students becoming hesitant to use reasoning during argument-based inquiry classes as students may feel pressured to respond to their teachers' questions with specified answers. Teachers should be encouraged to use open-ended questions rather than close-ended questions to promote their students' use of reasoning while learning new concepts.

Previous studies on teaching have shown that the academic achievement of students has a positively correlation with the number of clear academic questions generated by the teacher. Teachers' questions often have a structured and repetitive process. The process consists of (1) establishing the framework, background, or context of the question; (2) asking the question, and

(3) reacting to the question by correcting, clarifying, expanding, and praising (Clark, Gage, Marx, Peterson, Staybrook, and Winne, 1979). Woolfolk and McCune-Nicolich (1984) asserted that these questions might be asked corresponding to one of six levels of Bloom's cognitive taxonomy of objectives for student learning. Furthermore, Woolfolk and McCune-Nicolich (1984) suggested that diverse types of questions can be effective for student learning. The types of questions asked by the teacher should consider many factors; (1) the instructional objectives, (2) the student's age, (3) socioeconomic background, and (4) ability. This research indicated that teachers should not feel obligated to use open-ended questions all of the time. They may need to ask diverse types of questions considering students' cognitive development and background knowledge for student learning. They also suggested examples of using diverse types of questions depending on students' achievement level.

For low academic-achieving students, questions that ask for simple factual information/knowledge or questions that ask for understanding may be more successful, because those types of question require convergent responses (one correct answer) (Woolfolk and McCune-Nicolich, 1984). Thus, students are able to capture the meaning of the question easily. For these students, teachers should phrase questions that lead to correct responses frequently accompanied with praise and encouragement.

For high academic-achieving students, teachers should ask more high-quality (open-ended) questions that are divergent (multiple correct answers) (Woolfolk and McCune-Nicolich, 1984). For these high academic-achieving students, teachers should also reduce the frequency of encouragement, praise, and teacher-directed discussion in order to let the students talk as much as possible**.** Teacher-directed discussion takes place on a limited basis for the purpose of clarifying, correcting, or criticizing students' responses when it is necessary.

For the classes composed of a mix of academic-achieving students or students with low self-esteems, a mix of open-ended and closed-ended questions should be utilized alternatively, accompanied by proper criticism or praise and encouragement as deemed appropriate considering the instructional goals and needs of the student (Medley, 1977; Ward & Tikunoff, 1976).

Ramsey et al (1990) summarized the role of teacher in using questions in class as follows;

1. Ask knowledge-level questions when assessing students' ability to recall, recognize, or repeat information as it was learned.
2. When assessing students' higher-level thinking, use terms such as *how, why, what if* to encourage deeper thought.
3. Prepare questions in advance.
4. Ask questions in a logical sequence.
5. Ask specific questions that students can answer silently.
6. Ask direct questions are asked, sprinkle the questioning with direct statements.
7. Request that students repeat the teacher's question before answering.
8. When a specific student is asked a question, have another student repeat the question before allowing a response.
9. Allow students to converse with each other in a student-directed manner after a question is asked or answered.
10. Request that students express their own questions fully and specifically.
11. Name specific students in a random order to respond to questions.
12. Provide adequate waiting time after naming a respondent. Rowe (1974) found that waiting approximately three to five seconds after naming a respondent before eliciting a response brought better responses from more students (p. 421-422)

His suggestions for using questions in class were somewhat teacher-centered (e.g., suggestions 1, 3, 4, 5, 6, 7, 8 and 11) based on the classifications of traditional teacher questioning (close-ended questions) of table 2. 1. However, the core concept of open-ended (student-centered) (e.g., suggestions 2, 9, 10 and 12) based on the classifications of teacher constructivist/inquiry questioning (open-ended questions) of table 2. 1, was also reflected in his suggestions. It is difficult for teachers to use open-ended (student-centered) questions all the time because not every student possesses the same baseline when they learn new concepts. Teachers should be encouraged to use their questioning in terms of alternating between teacher-centered and student-centered depending on student cognitive levels (Chin, 2007; Scott, 1998). Ramsey's suggestions provide a balanced instruction between teacher-centered and student-centered questioning for teachers.

More recent research has focused on incorporating inquiry into teacher questioning so teacher questioning became student-centered. They have tried to link teacher questioning to student learning. Chin and Osborne (2008), Mortimer and Scott (2003), and King (1996) elucidated the role of teachers in inquiry-based classes by emphasizing the fact that the core of teaching science is to connect students to the social language of school science. The teacher must introduce scientific ideas to students within a social context and encourage students to develop the ability to apply the ideas to a variety of situations. In doing so, the students are eventually able to internalize new ideas and make sense of the new ideas.

The role of teachers in inquiry-based classes can be summarized as promoting the use of open-ended questions. Open-ended questions play a role in eliciting student reasoning processes. In the long run, students will be able to benefit from teachers' open-ended questions. It is difficult for teachers to shift from traditional questioning (close-ended questions) to inquiry-based questioning (open-ended questions) as previous research revealed (Chin, 2007; Oliveira, 2010; Scott, 1996) even though teachers understand the effect and importance of utilizing open-ended questions in classes. However, studies about the effect of open-ended questions on student learning in terms of facilitating students to use reasoning while learning new concepts has been conducted and evaluated by many researchers. Results have indicated that teachers should be encouraged to use open-ended questions more than close-ended questions to promote student learning.

The effect of teacher questions on student reasoning during the learning process

Results of research regarding teacher questioning that facilitates student reasoning

Teachers in the classroom very frequently ask numerous follow-up questions to students before any response can be given from the students (Ramsey et al, 1990). However, the "quantity of questions asked does not necessarily demand quality responses on the part of the learner" (Dean, 1986, p. 184). Follow-up questions should be generated by the teacher that align with student cognitive steps from lower-order to higher-order thinking, and to "usher the habit of an intellectual pause in the discussion, refreshing both leaders and participants" (Will 1987, p. 34). Effective questions can be generated to lay the foundation for some of this scaffolding for student learning. For example, by requiring students to clarify, elaborate, or justify their ideas in their responses (Walsh & Sattes, 2005). Teacher questioning as a scaffolding strategy should correspond to student responses (Lee & Kinzie, 2012). A number of studies have found a positive impact of teachers' open-ended questions on student learning, particularly in language achievement (e.g., Chin, 2007; Chin & Osborne, 2008; Conezio and French 2002, de Rivera et al. 2005; Lee & Kinzie, 2012; Wasilk et al. 2006; Whitehurst et al. 1994).

Additionally, McNeill and Pimentel (2010) and Martin and Hand (2009) supported the effect of open-ended questions in argument-based inquiry classes by suggesting that there was a relation between a higher frequency of open-ended questions from teachers, a higher percentage of student talk, a higher frequency of using reasoning in terms of providing evidence to support their claims, and a higher interaction between students in terms of argumentation. In other words, teachers did not talk most of the time. Students had opportunities to talk about ideas that were discussed during the classes, utilizing argumentation components.

Based on the studies about the effect of teacher questioning types on student responses (student learning), two flow charts for the effects of open-ended and close-ended questions on student learning were generated by the researcher. The two flow charts depict the analytical framework for the teacher questioning and student response categories that have emerged.

Figure 2. 1. Effect of open-ended questions

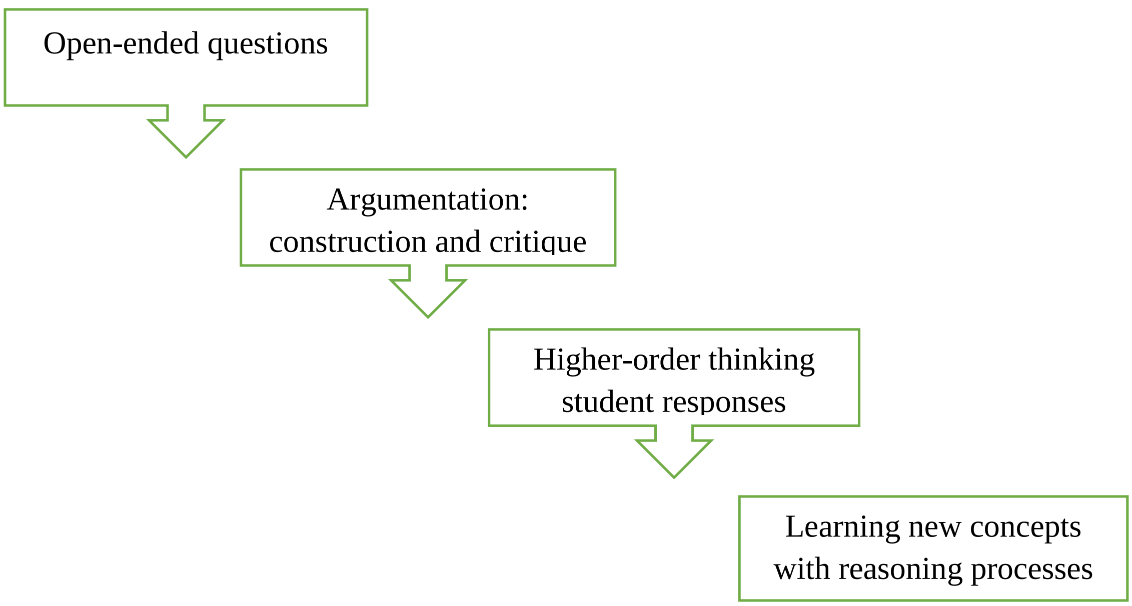


Figure 2. 1 conceptualizes the effect of open-ended questions on student learning in terms of higher-order thinking responses. Open-ended questions elicit argumentation components from students and that leads to higher-order thinking responses while they are learning new concepts (Zee & Minstrell, 1997b). Open-ended questions containing argument components and Bloom's revised higher level taxonomy components (challenging, defending, supporting, persuading, and rejecting/rebuttal with evidence) can be divided into three sub-types of open-ended questions: (1) asking for explanation (AE), (2) asking for self-evaluation of reasoning (AF), and (3) asking self-

evaluation of others' reasoning (AFO). These sub-types of open-ended questions were used for the analytical framework for the teacher questioning category. Higher-order thinking student responses in relation to teachers' sub-types of open-ended questions were divided into three sub-types: (1) explanation responses (E), (2) self-evaluation of reasoning responses (SE), and (3) self-evaluation of others' reasoning responses (SEO). These sub-types of higher-order thinking student responses were used for the analytical framework for the student response category. Through this process, students are able to respond to the questions using higher-order thinking on a more frequent basis. More detailed explanations of the framework is provided at the end of the chapter.

The degree of reasoning in student responses corresponds to Bloom's revised taxonomy. Open-ended questions correlate with higher-order thinking student responses (Martin & Hand, 2009; McNeill & Pimentel; 2009; Zee & Minstrell, 1997b).

Table 2. 2. Bloom's revised taxonomy for the degree of reasoning in student responses

Type / Cognitive level	Bloom's revised taxonomy
Higher order thinking	Creating
	Evaluating
	Analyzing
	Applying
Lower order thinking	Understanding
	Remembering

The components in table 2. 2 relate to the argumentation components (Duschl et al, 2007; McNeill & Pimentel, 2009; Osborne et al, 2004; Toulmin, 1958). The argumentation components consist of claim, challenge, defend, support, rebuttal/reject, and persuade ideas with evidence. Table 2. 3 explains the relations between the Bloom's revised taxonomy components and the argumentation components in detail.

Table 2. 3. Relating Bloom's revised taxonomy to argumentation components

	Taxonomy	Argumentation
Higher-order thinking	Creating	Claim, challenge, defend, support, rebuttal and persuade with evidence
	Evaluating	Claim, challenge, defend, support, and rebuttal with evidence
	Analyzing	Claim, challenge, support, and defend with evidence
	Applying	Claim, challenge, and defend with evidence

Table 2. 3 shows how open-ended questions elicit higher-order thinking student responses by relating argumentation components to student responses with higher-order thinking, which corresponds to Bloom's revised taxonomy components (Berland & Reiser, 2009; McNeill & Pimentel, 2009; NRC, 2008; Suppe, 1998). The lowest rank of the higher-order thinking student responses in the revised taxonomy is applying ideas. Student responses using applying ideas can be achieved by using three argumentation components. Students respond by making claims regarding their ideas first, their claims are challenged by their peers and/or teachers and defended by themselves. The next component is analyzing ideas. Students go through four steps. They make claims regarding their ideas. Their claims are challenged by others. Their claims are defended by themselves. Their claims are supported by their peers and/or teachers. The next component is evaluating ideas. Students make claims. Their claims are challenged by their peers and/or teachers. They defend their claims by providing evidence. Their peers support or reject their claims by providing evidence. The highest component is creating ideas. Students make claims on their ideas. Their claims are challenged by their peers and/or teachers. The students defend their claims by providing evidence. Their claims are supported and rebutted by their peers and/or teachers with evidence. Through this process, everyone in class reaches a consensus and this leads to creating a new claim. Argumentation is a core component in language when students

learn new concepts (Kim & Hand, 2015). Argumentation can apply to any subject since teaching takes place in terms of language (e.g., representations, writing, and verbal expressions).

Open-ended questions also enable teachers to create a learning environment where students are encouraged to talk during class utilizing argumentation components (Akkus, Gunel, & Hand, 2007). Creating a supportive learning environment where open-ended questions are used for students is necessary in order to provide students time to think about topics and to engage in the activity (Fahy, 2004).

In contrast, research suggests that close-ended questions may inhibit student learning. The function of close-ended questions is to look for factual information, promote the recall of information, or to confirm information without asking for reasoning while learning new concepts (Blosser, 1973; McNeill & Pimentel, 2009). In other words, close-ended questions look for correct answers not diverse answers for questions.

Figure 2. 2. The effect of close-ended questions

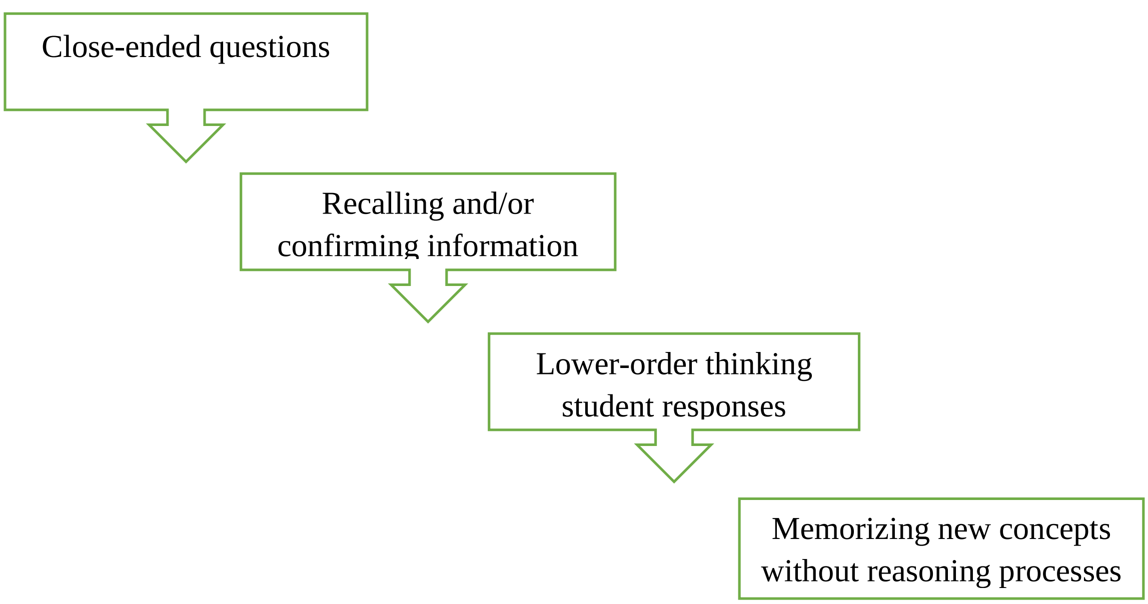


Figure 2. 2 conceptualizes the effect of close-ended questions on student learning. Close-ended questions ask for factual information or confirmation. This leads to memorizing new concepts using lower-order thinking processes rather than understanding them via higher-order thinking (reasoning processes). Lower-order thinking student responses in table 2. 2 correlate with close-ended questions. Close-ended questions that do not contain argument components and Bloom's revised higher-level components (Blosser, 1973; McNeill & Pimentel, 2009) can be divided into two sub-types of close-ended questions: (1) asking for factual information (AI), and (2) asking for confirmation (AC). These sub-types of close-ended questions were used for the analytical framework for the teacher questioning category. Lower-order thinking student responses in relation to teachers' sub-types of close-ended questions were considered as one sub-type: simple responses (S). This sub-type of lower-order thinking student responses were used for the analytical framework for the student response category. Detailed explanations of the analytical framework is provided at the end of the chapter.

Methods to facilitate teacher questions appropriately to enhance student learning

The IRE or IRF sequence focuses on delivering content knowledge rather than facilitating student reasoning while students are learning new concepts. In order to resolve the problem of the IRE or IRF sequence, Mortimer and Scott (2003) redeveloped the IRE or IRF sequence by expanding it into the IRFRF model in which a further answer from the student follows the expounded feedback from the teacher. This commonly occurs in dialogic discourse. During the feedback, the teacher uses a student's comment or idea to encourage the student to continue, elaborate on the comment or idea, or provide elaboration.

NRC (2012) also suggested how to use questions that can elicit student reasoning as follows;

(1) Learners are engaged in scientifically oriented questions.

(2) Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions.

(3) Learners formulate explanations from evidence to address scientifically oriented questions

(4) Learners evaluate their explanations in light of alternative explanations, particularly those reflecting scientific understanding

(5) Learners communicate and justify their proposed explanation.

In summary, the practices described above play out in the following steps: (a) drawing, challenging, and developing students' ideas (b) taking into account peer assessment in evaluating students' answers, (c) using the "reflective toss strategy", and (d) considering wait time.

Student responses in regard to teacher question types

Zee and Minstrell (1997b) suggested that when teachers used open-ended questions frequently, student responses were accompanied by student reasoning processes. Martin and Hand (2009) evaluated the relation between teacher question types (strategies) and student engagement (student response types). They discovered that teacher question types were directly related to student response types. The participating teacher changed their question types over two years and that led students to participate actively in the classes and to provide explanations using reasoning based on evidence. Kim and Hand (2015) emphasized that types of teacher discourse (e.g., being non-directional in terms of challenging ideas) affected student response types (e.g., challenging, supporting, defending, and rejecting with evidence). Student responses or student talk is an important factor when students learn new concepts (Barnes, 1976; Britton, 1982; Bruner, 1986; Corson, 1988; Lemke, 1990; Maloch, 2002; Pea, 1993; Prawat, 1993; Rivard & Straw, 2000; Schoenfeld, 1989). Teacher questioning should encourage students to respond with higher-order thinking to promote reasoning in the learning process.

New directions for researching teacher questioning

There are limited empirical studies on the impact of open-ended questioning on students' cognition and learning (Lee & Kinzie, 2012). Furthermore, even though many researchers generally agreed that teachers' questions aimed at facilitating high cognitive levels tended to generate students' responses requiring higher-order thinking, some researchers disagreed with this finding by asserting that there is a low relation between the cognitive level of teachers' questions and the cognitive level of students' answers (e.g., Dillon 1982b; Mills et al. 1980). Because there is disagreement between researchers on the effect of open-ended questions on leading to higher-order thinking student responses, this study focused on finding patterns among student talk time, structure of questions, and student responses between the treatment (argument-based inquiry classes) and control groups (traditional classes). Kim and Hand (2015) suggested that teacher questions should be investigated to elicit student reasoning in an argument-based inquiry class based on their analysis of teacher and student discourse patterns in between argument-based inquiry classes and traditional classes. Their analysis results showed that the argument-based inquiry class teachers and students used critique components (challenge, defense, support and reject ideas), which are critical for argumentation, more frequently than the traditional class teachers and students. These results implied that teacher question types were related to student responses while learning new concepts. Furthermore, based on the results from the initial analyses, any relations among teacher talk time, structure of question, and student response will be investigated.

Emerging Analytical Framework for the study

Three categories emerged for the analytical framework for the study: (1) teacher talk time, (2) teacher question types, and (3) student responses. These three categories are related when students utilize inquiry to learn new concepts.

Teacher talk time

The first category was teacher talk time. Teacher talk time was originated based on the work of Akkus, Gunel, and Hand (2007), Fahy (2004), Kim and Hand (2015), and Rowe (1974). Rowe (1974) suggested that teachers should provide enough wait time for student responses after teachers ask questions. Kim and Hand (2015) investigated the duration of teacher talk time in terms of directionality between argument-based inquiry classes and traditional classes. Teachers in argument-based inquiry classes talked less and encouraged their students to talk more by using argumentation components compared to teachers in traditional classes. Teachers in traditional classes provided directions when students got confused about the concepts they were learning by talking frequently. In contrast, teachers in argument-based inquiry classes led students to explore the concepts on their own by using argumentation components in their discourses. Akkus, Gunel, and Hand (2007), Fahy (2004) and NRC (2000, 2012) also suggested that in order for students to engage in inquiry, teachers should be encouraged to create a supportive learning environment where students feel comfortable with expressing their ideas without feeling pressured to respond with correct answers.

Teacher question types

Among the questions types, open-ended and close-ended question types were adopted for the study. Open-ended questions include any type of question that promotes student reasoning when they learn new concepts. Close-ended questions include any type of question that do not promote student reasoning. Some of the question types (e.g., rhetorical questions and managerial questions) that were studied were not related to these categories (open-ended and close-ended question types) because those question types do not elicit student reasoning processes.

Under the open-ended question types, three sub types of open-ended questions emerged based on previous research: (1) Asking for explanation (AE), (2) Asking for self-evaluation of reasoning (AF), and (3) asking for self-evaluation of others' reasoning (AFO). The first sub type of open-ended question was originated from the work of NRC (2000) and Zembal-Saul, McNeill and Hershberger (2013). They emphasized the importance of explanation in the learning process. The second and third sub-types of open-ended questions were originated from Berland & Reiser (2009), Newton, Driver & Osborne (1999), Driver, Newton, & Osborne, (2000), and NRC (2000, 2008). They emphasized the importance of evaluating ideas in the learning process.

Under the close-ended questions, two sub-types of close-ended questions emerged based on previous research: (1) asking for factual information (AI) and (2) asking for confirmation (2). These two sub-types of close-ended questions were originated from Kearsley (1976), Long & Sato (1983), O' Connor and Michaels (1993), Oliveira (2010) and Waring (2002). They sub-divided close-ended question type into types of question that requires students to provide formation or definitions of concepts that they are learning without going through reasoning processes (e.g. reiteration and memorization) and to respond in terms of confirming what they understand without going through reasoning processes in a simple way (e.g., "do you understand?").

Student responses

Two types of student responses emerged: (1) higher-order thinking student responses and (2) lower-thinking student responses. Under the higher-order thinking student responses, three sub-types of student responses were classified: (1) explanation responses (E), self-evaluation of reasoning responses (SE), and (3) self-evaluation of others' reasoning responses (SEO). Sub-types of higher-order thinking student responses were originated based on the work of Anderson et al. (2001), Bell and Linn (2000), Dillon (1982a,b), Kim and Hand (2015), Lee and Kinzie (2012), Mcneill and Pimentel (2009), Mills et al. (1980), NRC (2012), and Toulmin (1958). They suggested students should be encouraged to respond with explanations and/or self-evaluations of one's own reasoning or others' reasoning when they learn new concepts.

Under the lower-order thinking student responses, only one sub-type of student responses emerged: simple responses (S). Simple responses were originated based on the work of Bloom's revised taxonomy. This type of response does not require students to use their reasoning while they answer their teachers' questions (e.g. remembering of ideas). Many researchers argued that using types of questions that elicit this type of student responses frequently may not be helpful for student learning (e.g., Kim and Hand, 2015, Lee and Kinzie, 2012; NRC, 2000; 2012).

The analytical framework for the study emerged based on previous work as explained above. A numerous studies supported that open-ended questions affect student learning positively in terms of promoting higher-order thinking student responses frequently when enough student talk time is provided. In other words, when teacher talk time is dominant in class, students do not have opportunities to express their ideas using inquiry components.

CHAPTER 3

METHOD

This chapter describes the research design, data collection procedures, and the process of analysis used in the study. The first part of this chapter explains the research design of the study and the rationale for the method in the study. The second part of this chapter describes the analysis procedures: (1) the context of the study, (2) data collection procedures, (3) steps of descriptive analysis with theoretical foundations and an explanation of each step, (4) rationale for the coding system, (5) codes for types of teacher questions and student responses, and (6) phases for statistical analysis (repeated measures ANOVA, and Cohen's d).

Research design

The main purpose of this study was to explore any difference, similarity or pattern of types of questioning (structure of questions) that impacts student learning by observing student responses in two different class settings with regard to teacher talk time. Two groups were compared: A control (traditional classes) group, and a treatment (inquiry-based classes) group. In order to achieve the purpose of the study, it was necessary to analyze teacher talk time, types of teacher questions, and student responses.

Quantitative research

The methods of investigating teacher questioning in two different group settings (inquiry-based class and traditional class) align with the intent of correlational quantitative research. This type of research is conducted to analyze and interpret data statistically to find relations between two or more variables to identify patterns or trends in the data. The researcher interpreted and evaluated the characteristics of teacher questioning (question types) in relation to teacher talk time and student responses in the context of diverse classrooms as well as within each classroom during the overall period of the project and at each time point. Furthermore, the patterns that teacher questioning had in relation to teacher talk time and student responses in different types of classrooms as well as within each group both over the period of the project and at each time point were evaluated statistically using repeated measures ANOVA, and Cohen's d for effect sizes. Through these statistical comparisons across settings as well as within each setting over the period of the project, the study examined the relations among types of teacher questions, teacher talk time and student responses.

Analysis procedures

This second part describes the analysis procedures that were used in the study. These include (1) the context of the study, (2) data collection procedures, (3) steps of descriptive analysis with theoretical foundations and an explanation of each step, (4) rationale for the coding system, (5) codes for types of teacher questions, and (6) phases for statistical analysis.

Context of the study

Teachers in this study participated in a randomized field trial of the Science Writing Heuristic [SWH] project in a Midwestern state for two years and were divided into a treatment group and a control group. Before the schools were assigned to one of the two groups, 48 participating schools were randomized at the beginning stage to control for any bias that might arise during the selection of participating teacher classrooms. The study focused on teacher questioning and talk time in relation to student responses. Students were not the main focus of the study. However, student responses were included as part of the study to investigate the effect of teacher question types and talk time on student learning in terms of student response types. The random field trial project included three hundred twenty teachers and over six thousand students. The students ranged from third grade to fifth grade.

Hand et al (2002) stated that SWH "…is intended to help students construct understanding during practical work. Students are required to produce written explanations of the processes involved in the activity through completion of a template, with particular emphasis placed on claims, evidence and reflection" (p. 20). The teachers in the treatment group of the project received an intensive PD session each summer before their fall semester started and one short PD session during each semester. The overall focus of all of the PD sessions for the treatment group was to engage the participants in experiencing the SWH approach. The short PD sessions focused on listening to the teachers' issues and resolving them. In addition, teachers from the treatment group shared their experiences with each other. The teachers also had opportunities to design their plans for topics that they needed to teach during the semester.

The teachers in the control group of the project received their normal professional development activities designated by the school as part of each school's improvement plan. However, they did not receive the same type of SWH PD sessions as the treatment group. Each school district is responsible for implementing a continuing education plan for their teachers, including an emphasis on science. All control and treatment schools continued with the regularly scheduled PD sessions provided by their school district. Thus, both the SWH/treatment teachers and control teachers were involved in PD programs each year. A total of forty-eight schools and three hundred and twenty teachers participated in the project. The average class size across both groups was twenty-two students. All of the teachers in the project were assigned into the treatment group (24 schools) or the control group (24 schools) based on their school.

Data collection

All of the teachers in the project were asked to send a video clip of their class at the end of each semester (fall and spring) for two years (i.e., four video clips per teacher). However, not all teachers sent four video clips. Of the one hundred teachers who sent all four video clips of their classes to the investigator (out of a total of three hundred and twenty teachers in the project), thirty teachers were chosen randomly from each group to explore how their talk time, questioning, and student responses based on the teachers' question types changed throughout the project. The participants were chosen using the excel program to reduce selection effects from a pool that randomized the one hundred qualified participants for the study. This study used convenience sampling because the data was previously collected.

Steps for descriptive analysis

The descriptive analysis of this study consisted of four steps. Table 3. 1 explains the relation between the research questions and the steps of the analysis.

Table 3. 1. Conceptualization of the relation between research questions and steps of descriptive analysis.

Research questions	Corresponding steps
Is there a difference in the amount of teacher talk time between in argument-based inquiry classes and teachers in traditional classes?	• First step: Teacher talk time
Are there differences in question types and student responses between in argument-based inquiry classes and teachers in traditional classes?	• Second step: teacher question types • Third step: student responses
What are the differences in teacher questioning and student responses between teachers in argument-based inquiry classes and teachers in traditional classes? What is the relation between teacher talk time, teacher questioning and student responses within each group?	• Fourth step: synthesizing the previous three steps

Step 1

The first step was to evaluate sixty teachers' teacher talk time at the end of each semester (fall and spring) during two years to examine any changes in the duration of teacher talk time across both groups and student interaction with their peers or teachers throughout the project. Teacher talk time was defined as the amount of time (seconds) the teacher lectures, explains or gives directions by asking questions to students. Personal conversations with students were not included as talk time. Teacher talk time was measured in seconds. Altogether, two hundred and forty video clips were used to collect data.

Step 2

The second step was to investigate teacher question types across the treatment group and the control group. The types of teacher questions were coded based on Bloom's revised taxonomy by Anderson et al (2001) and Blosser's (1973) four types of questions. Anderson and her colleagues made two changes from the original cognitive domain in Bloom's taxonomy: (a) they modified the form of the words from nouns to verbs, and (b) they reorganized the top two words in the taxonomy by changing the order (see Table 3. 2) (Anderson, et al, 2001; Pohl, 2000). The revised table reflected a more accurate explanation of the cognitive level of the learners that the researcher intended to evaluate than the original table in terms of relating argumentation components to higher-order thinking components (see table 2. 3).

Table 3. 2. Comparison of Bloom's taxonomy and revised Bloom's taxonomy.

Type / Cognitive level	Bloom's taxonomy	Bloom's revised taxonomy
Higher order thinking	Evaluation	Creating
	Synthesis	Evaluating
	Analysis	Analyzing
	Application	Applying
Lower order thinking	Comprehension	Understanding
	Knowledge	Remembering

Blosser (1973) classified four types of questions: (1) open, (2) closed, (3) rhetorical, and (4) managerial. However, the structure of question category involved only open and closed-ended questions, which corresponded to the cognition level of the revised Bloom's taxonomy and King's (1994) open-ended question type. Rhetorical and managerial questions were not included in this category because they were not related to student learning in terms of facilitating student responses. These types of questions focused on attracting students' attention and managing classrooms rather than eliciting students' reasoning processes. This category was designed to distinguish between questions that required higher-order (higher-cognitive) thinking and questions that required lower-order (lower-cognitive) thinking. Additionally, it is important for the teacher to prepare and ask a series of questions that considers the level of the learner's understanding of a concept. Mortimer and Scott (2003) created the Initiation-Response-Feedback-Response-Feedback [IRFRF] sequence. This enables the learner to construct their own comprehension through interacting with their teacher or their peers rather than being asked to memorize factual information without the reasoning process (Chin, 2007; Roth, 1996). Teacher questioning needs to be structured to promote student learning so the learner is able to reach the goals that were articulated by NRC (1996, 2000).

Step 3

The third step was to analyze and code student responses based on the revised Bloom's taxonomy since it was designed to assess the learner's cognitive level. The revised Bloom's taxonomy defined understanding and remembering as lower-order thinking processes since remembering does not require the reasoning process from the learner. Understanding concepts requires learners to use the reasoning process to a certain degree but not to the same level as higher-order thinking processes. Understanding requires two roles: (1) the role of learner who follows the steps that are suggested by teachers, and (2) the role of teacher who suggests steps for the learner. In other words, learners do not have an opportunity to utilize their reasoning process fully. Thus, the degree of utilizing students' reasoning processes is directed by the role of teacher compared to higher-order thinking processes. However, when the learner's answer includes applying, analyzing, evaluating, and/or creating knowledge, it is defined as a higher order-thinking process since those components required the learner to use the reasoning process actively. Student response in NRC (1996, 2000) was defined in a similar way in Bloom's revised taxonomy (e.g., higher-order thinking and lower-order thinking).

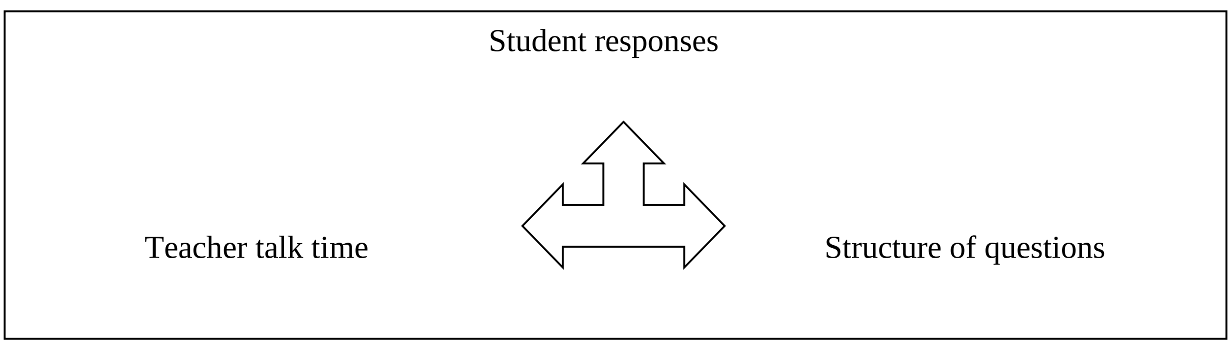
Step 4

The fourth step was to synthesize the previous three steps to find any relation between those components of teacher questioning in the two different groups over the period of the project (e.g., any difference, similarity or pattern regarding teacher talk time, structure of questions, and student responses).

Rationale of coding system

The above steps were conducted to evaluate patterns and relations between those three categories in both groups to answer the research questions of the study. The conceptualization of the coding system is shown in figure 3. 1.

Figure 3. 1. Conceptualization of coding system



The coding system conceptualized the relations between the three categories in both groups. Teacher talk time interacted with the structure of questions (question types) and student responses. The structure of questions interacted with teacher talk time and student responses. Student responses interacted with teacher talk time and structure of questions.

Teacher talk time was measured to evaluate any difference between the control and treatment groups, at each time point between the groups, and at each time point within each group at the starting stage of the study. Based on the results, the relations between teacher talk time and structure of questions, or student responses were investigated (e.g., Carlsen, 1998, 1991; Chin, 2007; Mortimer & Scott, 2003; NRC, 1996, 2000; Roth, 1996). The next step was conducted to evaluate the interactions between the structure of questions and student responses.

Through these processes, patterns and relations among the three categories between groups, at each time point between groups and at each time point within each group were measured. For these reasons, the four steps of the coding system were developed for the purpose of the study.

Codes for types of teacher questions

In order to analyze teacher questioning, two main categories were measured. The first category was structure of teacher questions. The second category was student response. Each category is explained with resources and examples in tables 3. 3, 3. 4, 3. 5, and 3. 6.

Structure of questions (question types)

This category was intended to evaluate the type of teacher question in both groups**.** Table 3. 3 explains the resources of this category.

Table 3. 3. Resources of question structure

Category	Corresponding resources	Used components
Structure of questions • Open-ended questions • Close-ended questions	• Bloom's revised taxonomy (Anderson, et al, 2001) • Question types (Blosser, 1973; King, 1994)	• All of the components • Two question types (open-ended and close-ended questions)

This category was developed based on two theories: Bloom's revised taxonomy and question types. In the Bloom's revised taxonomy, all of the six components were used. In the question types, two question types (open-ended and close-ended question types which facilitate student reasoning processes) out of four types were used. Rhetorical and managerial questions, were not used because rhetorical questions focused on making an effect or an assertion about ideas for students to engage in the ideas rather than eliciting an answer and managerial questions are aimed at classroom managing. The purposes of the two question types do not lead to promoting student reasoning processes while learning concepts. The following table explains the codes for question types with descriptions and examples.

Table 3. 4. Structure of questions

Type	Description	Code	Description for code	Example
Open-ended question (**O**)	Elicits student reasoning (Reasoning is defined as the process of leading to inferences/explanations through thinking based on facts/information)	Asking for explanation based on experience, evidence or data (**AE**)	Question requires explanation based on experience or data (Explanation is defined as " attempt to provide an account that specifies what happened and/or why it occurred")	How does air gets into you compared to your model?
		Asking for self-evaluation of reasoning (**AF**)	Question requires self-evaluation of one's own idea	How did you make it different or better?
		Asking for evaluation of other's reasoning (**AFO**)	Question requires self-evaluation of others' ideas	What do you think (about his idea)?
Closed-ended question (**C**)	Elicits factual information or confirmation	Asking for factual information (**AI**)	Question requires factual information without reasoning	When we do research, we need an expert's help. Does that help?
		Asking for confirmation (**AC**)	Question requires confirmation without reasoning	Does your model work the way the human body works?

This category has two question types: (1) open-ended question and (2) closed-ended question. Question types were categorized based on the revised Bloom's taxonomy and two of Blosser's (1973) four question types. Open-ended questions were used to see how questions helped students elicit higher higher-order (higher-cognitive) thinking. In contrast, close-ended questions were used to see what type of questions produced reiteration of factual information or elaborating knowledge without the reasoning process on the student's part.

Open-ended question (O) consisted of three types of questions: (1) asking for explanation based on experience/data, (2) asking for self-evaluation of reasoning, and (3) asking for evaluation of other's reasoning.

The first open-ended question was *asking for explanation based on experience/da*ta. It was defined as asking students to explain their or others' thinking based on experience or data by comparing or contrasting their or others' ideas. An explanation was defined as "attempt to provide an account that specifies what happened and/or why it occurred" (Berland & Reiser, 2009, p. 27). For example, "How does air get into you compared to your model?"

The second open-ended question was *asking for self-evaluation of reasoning*. It was defined as asking students to justify or evaluate their own thinking. Reasoning was defined as the process of leading to inferences or explanations through thinking based on facts or information. For example, "How did you make it different/better?"

The last open-ended question type was *asking for evaluation of other's reasoning*. It was defined as asking students to evaluate other students' thinking. For example, "What do you think (about his idea)? What else do you think? Any questions for them?"

Close-ended question (C) was comprised of two types of questions: (1) asking for factual information, and (2) asking for confirmation.

The first question type was *asking for factual information*. It was defined as asking for simple knowledge, experience, facts, elaborations or definitions without requiring any reasoning process. For example, "When we do research, we need an expert's help. Does that help?"

The second question type was *asking for confirmation*. It was defined as teacher's questions that asked students for agreement, disagreement, endorsement or understanding of ideas, knowledge, explanations, definitions or facts without requiring any logical process. For example, "That (your model) is representing the human body? Is that what will happen? Does your model work the way the human body works?"

Student response

This category was intended to code student responses in relation to the level of cognition. Table 3-5 explains the resources of student responses.

Table 3. 5. Resources of student response

Category	Corresponding resources	Used components
Student response(s) • Lower-order thinking • Higher-order thinking	• Bloom's revised taxonomy (Anderson, et al, 2001) • Student response (NRC, 1996, 2000)	• All of the components • Types of student response (higher and lower-order thinking)

This category was developed based on two theories: Bloom's revised taxonomy and the Five-E model. In the Bloom's revised taxonomy, all of the six components were used. In the student response, types of question (higher-order thinking and lower-order thinking) were used. The following table explains the codes for student responses with descriptions and examples.

Table 3. 6. Student response

Type	Description	Code	Description for code	Example
Lower-order thinking (**L**)	Student response shows memorizing or/and understanding an idea without the reasoning process	Simple response (**S**)	Responding yes or no, elaborating or reiterating information, facts, knowledge or experiences, or asking for information or confirmation.	Have you seen models? Yes.
Higher-order thinking (**H**)	Student response shows explaining, applying, analyzing, evaluating or/and justifying one's or/and others' ideas or leading student response to modifying one's or others' ideas by going through either the process of explaining, applying, analyzing, evaluating or/and justifying	Explanation response (**E**)	Responding with explanation	How did the air get into the lungs? When we breathe in, air comes into us.
		Self-evaluation of one's own idea response (**SE**)	Assessing one's own reasoning	Does your model work the way the human body works? No, it didn't work as I expected, I missed some parts in my model.
		Self-evaluation of others' ideas response (**SEO**)	Assessing others' reasoning	Why did you do that?

It consisted of two student responses: (1) lower-order thinking, and (2) higher-order thinking.

Lower-order thinking (L) was defined as a student response that showed memorizing and/or understanding an idea without the reasoning process. This type of student response was comprised of a simple response. Simple response indicated that students (a) responded in the form of “yes” or “no” either verbally or nonverbally, such as raising hands instead of responding "yes" or "no" to the teachers’ questions verbally, (b) elaborated or reiterated information, experience, facts or definitions, or (c) asked for clarification, confirmation or information for others’ ideas. For example, “Have you seen models? Yes”.

Higher-order thinking (H) was defined as a student response that showed explaining, applying, analyzing, evaluating and/or justifying one’s and/or others’ ideas or leading student response to modifying one’s or others’ ideas by going through either the process of explaining, applying, analyzing, evaluating and/or justifying. This type of student response consisted of three types: (1) Explanation response was defined as a student’s response that provided explanations in response to questions. For example, “How did the air get into the lungs? When we breathe in, air comes into us”, (2) self-evaluation of one’s own idea response, and (3) self-evaluation of others’ ideas response.

Self-evaluation of one’s own idea response was defined as a student’s assessment of their reasoning with or without experience or data and/or coming up with a modification of their reasoning. For example, “Does your model work the way the human body works? No, it didn’t work as I expected, I missed some parts in my model”.

Self-evaluation of others’ ideas response was defined as a student’s assessment of others’ reasoning with or without experience or data and/or coming up with a modification of others’ reasoning. For example, “Why did you do that?”

Phases for statistical analysis

The sixty teachers’ video clips were analyzed, coded and interpreted by the researcher as the basis of a statistical analysis (repeated measures ANOVA) along with effect sizes. The study investigated characteristics of teacher questioning across three categories: (1) teacher talk time (2) Structure of questions, and (3) student responses. The first category measured teacher talk time in seconds in both groups to investigate any relation between teacher talk time and characteristics of teacher questioning, and student responses. The second category (structure of questions) evaluated whether questions promoted student reasoning**.** In this category, the number of open-ended questions and close-ended questions was counted to compare the difference between teachers in the argument-based inquiry classes and traditional classes. The third category (student responses) assessed whether student responses demonstrated higher-order reasoning. In this category, the number of lower-order reasoning and higher-order reasoning responses were counted. The definition and an example of each category of codes and the analysis process are provided in detail in the following section.

Based on the number of codes from the three categories: (1) teacher talk time, (2) types of question, and (3) student responses, a repeated measures ANOVA was conducted to find the patterns and relations of their talk time, question types, and student responses throughout the project in both groups (e.g., difference or similarity in teacher questioning patterns in two diverse groups). The following table specifies the phases of the statistical analysis. The null and alternative hypotheses for the study are provided after the table 3. 7.

Table 3. 7. Statistical analysis phases

Phases	Data	Statistical tests
1. Teacher talk time	Teacher talk time measured in seconds and converted into percentages	Repeated ANOVA • Between groups • At each time point between groups • At each time point within each group Effect size • At each time point between groups
2. Question types	Open-ended question **(O)** • Asking for explanation based on experience, evidence or data **(AE)** • Asking for self-evaluation of reasoning **(AF)** • Asking for evaluation of others' reasoning **(AFO)** Close-ended question **(C)** • Asking for factual information **(AI)** • Asking for confirmation **(AC)**	Repeated ANOVA • Between groups • At each time point between groups • At each time point within each group Effect size • At each time point between groups
3. Student responses	Lower-order thinking **(L)** • Simple response **(S)** Higher-order thinking **(H)** • Explanation response **(E)** • Self-evaluation of one's own idea response **(SE)** • Self-evaluation of others' ideas response **(SEO)**	Repeated ANOVA • Between groups • At each time point between groups • At each time point within each group Effect size • At each time point between groups
4. Relation between talk time, question types, and student responses	Talk time, question types, and student responses	Repeated ANOVA and Effect size • Within each group

The data was analyzed in four phases.

Phase 1

Phase 1 measured teacher talk time (1) between the treatment and the control groups, (2) between groups at each time point, and (3) within each group at each time point. Teacher talk time was measured in seconds. The seconds were converted into percentage of class time to control for differences between class durations. A repeated measures ANOVA was used to evaluate the difference between the treatment and control groups in teacher talk time. Changes in teacher talk time within each group and between the groups at each time point were analyzed by a repeated measures ANOVA.

Null hypotheses and alternative hypotheses

In phase 1, three null hypotheses and three alternative hypotheses were tested.

The first null hypothesis was that there was no difference between the treatment and the control groups in the percentage of teacher talk time.

- The alternative hypothesis was that the treatment group teacher talk time was different from the control group teacher talk time.

The second null hypothesis was that there was no difference between the treatment and the control groups in teacher talk time at each time point.

- The alternative hypothesis was that there was a difference between the treatment and the control groups in teacher talk time at each time point.

The last null hypothesis was that there was no difference within each group in teacher talk time at each time point.

- The alternative hypothesis was that there was a difference within each group in teacher talk time at each time point.

Phase 2

In phase 2, the types of teacher questions were analyzed. There were two question types (open-ended and close-ended questions). Each question type was divided into several sub-question types. The open-ended question had three sub-questions: asking for explanation (AE), asking for self-evaluation of reasoning (AF), and asking for evaluation of others' reasoning (AFO). The close-ended question had two sub-question types: asking for factual information (AI) and asking for confirmation (AC). The frequency of each question type and sub-types were counted and converted into percentages to control for the difference in class time duration among teachers in both groups. In this phase, teacher question types (1) between the treatment and control groups, (2) between the groups at each time point, and (3) within each group at each time point were analyzed to find if there was a statistical difference using repeated measures ANOVA.

Null hypotheses and alternative hypotheses

In phase 2, three null hypotheses and three alternative hypotheses were tested to examine any difference in teacher question types.

The first null hypothesis was that there was no difference in teacher question types between the treatment and the control groups.

- The alternative hypothesis was that there was a difference in teacher question types between the treatment and the control groups.

The second null hypothesis was that there was no difference in teacher question types between the groups at each time point.

- The alternative hypothesis was that there was a difference in teacher question types between the groups at each time point.

The last null hypothesis was that there was no difference in teacher question types within each group across the time.

- The alternative hypothesis was that there was a difference in teacher question types within each group at each time point.

Phase 3

Phase 3 was conducted to examine student responses. There were two student response types (lower-order thinking and higher-order thinking responses). The lower-order thinking student response has one sub-response (simple response (S)). Student response types were divided into three sub-types (explanation response (SE), self-evaluation of one's own idea response (SE) and self-evaluation of others' ideas response (SEO)). The frequency of each response type and sub-types was converted into percentages to control for the differences in class time duration among teachers in both groups. The main purpose of this phase was to identify if there was a difference in student responses (1) between the treatment and the control groups, (2) between the groups at each time point, and (3) within each group at each time point using repeated measures ANOVA.

Null hypotheses and alternative hypotheses

Three null hypotheses and three alternative hypotheses were tested to examine any difference in student responses.

The first null hypothesis was that there was no difference in student responses between the treatment and the control groups.

- The alternative hypothesis was that there was a difference in student responses between the treatment and the control groups.

The second null hypothesis was that there was no difference in student responses between the groups at each time point.

- The alternative hypothesis was that there was a difference in student responses between the groups at each time point.

The last null hypothesis was that there was no difference in student responses within each group at each time point.

- The alternative hypothesis was that there was a difference in student responses within each group at each time point.

Phase 4

In phase 4, the relations between the three categories (teacher talk time, teacher question types, and student responses) within each group were analyzed. In order to find relations between the three categories in each group, relations between the three categories were identified based on the results of repeated measures ANOVA and effect sizes

The effect sizes for the three categories were calculated in order to find the degree of the effect of the SWH approach for three categories (teacher talk time, structure of questions, and student responses). Cohen’s d was used to calculate effect sizes by using the means and standard deviations from two groups (treatment and control). Cohen’s d is a scale free and descriptive measure of characteristics or categories between groups. In other words, the magnitude of effect of the intervention between groups in terms of characteristics or categories can be calculated regardless of different units. Cohen’s d is easy to understand (see table 3. 8) and is used widely in numerous research studies. It only needs means and standard deviations of groups to be compared. The equation is as follows:

$$\text{Cohen's } d = M_1 - M_2 / s_{pooled}$$

$$\text{where } s_{pooled} = \sqrt{[(s_1^2 + s_2^2) / 2]}$$

M: mean of a group, s: standard deviation of a group

Cohen provided a guideline for the degree of effect of the intervention.

Table 3. 8. Effect size guidelines

Degree of effect size	Small	Medium	Large
Range	0.2 – 0.5	0.5 – 0.8	> 0.8

When an effect size is greater than 0.2 and less than 0.5, it is considered as small. When an effect size is greater than 0.5 and less than 0.8, it is considered as medium. When an effect size is greater than 0.8, it is considered as large. The effect sizes for the three categories between the treatment and control groups were presented in chapter 4.

Summary

Two randomly assigned groups (SWH group and traditional group) out of forty- eight schools participated in the study for two years. There were three categories in the analytical framework for the study: (1) teacher talk time, (2) the structure of questions, and (3) student responses. Teacher talk time was measured in seconds. The frequency of the structure of questions and student responses was counted and converted into percentages to control for the differences in class duration. Two statistical methods: (1) Repeated measures ANOVA and (2) Cohen's d, were used to find any differences, relations and the magnitude of effect of the SWH approach among teacher talk time, the structure of question (question types) and student responses between the two groups (treatment and control groups), between the two groups at each time point and within each group at each time point. There were four phases to answer the research questions for the study.

CHAPTER 4

RESULTS

This chapter consists of two parts. The first part provides descriptive analysis results and explains a general pattern of the three categories based on (1) the duration of teacher talk time and frequencies of the two categories: (2) question types and (3) student responses over the period of the project and at four different time points between the treatment and control groups and within each group. The second part presents interpretive analysis results. Based on the descriptive analysis results, statistical analyses were conducted to evaluate the statistical differences between the treatment and control groups and within each group over the period of the project and at each time point in terms of the three categories using repeated measures ANOVA with the alpha level .05 along with effect sizes. Relations between the three categories were identified based on the results of repeated measures ANOVA and effect sizes.

Descriptive analysis results

Overall results from each category were averaged across the experimental and control groups to compare the characteristics of teacher questioning. In each category, detailed analyses are presented. Teacher talk time was averaged across each group. Teacher talk time was defined as the amount of time (seconds) the teacher lectured, explained or gave directions by asking questions to students. Personal conversations with students were not included as talk time.

Teacher talk time

Table 4. 1. shows the difference in talk time between traditional teachers and argument-based inquiry teachers.

Table 4. 1. Average percentage of teacher talk time in traditional and argument-based inquiry classes

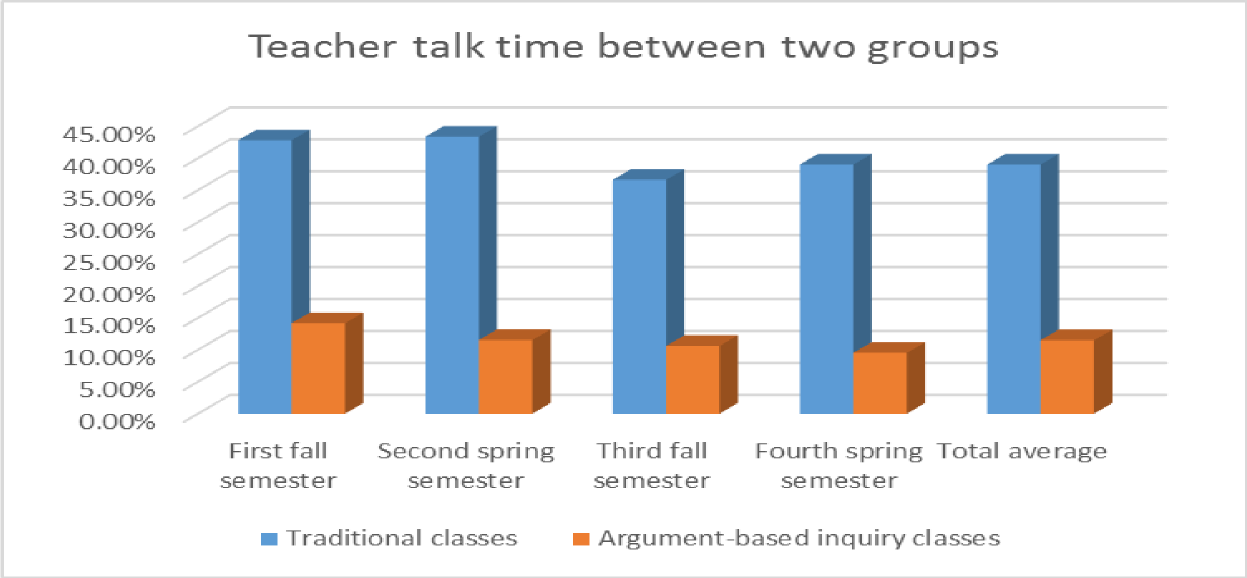
Group / Semester	Traditional classes	Argument-based inquiry classes
First year		
Fall semester	42.8%	14.17%
Spring semester	43.36%	11.53%
Second year		
Fall semester	36.60%	10.61%
Spring semester	39%	9.52%
Total average	39%	11.49%

There was a clear difference in teacher talk time between teachers in traditional and argument-based inquiry classes across the duration of the project time based on the percentage of teacher talk time between the two groups, between the two groups at each time point and within each group at each time point. Teachers in traditional classes spent more time talking during class compared to teachers in argument-based inquiry classes. The total average talk time across the teachers in traditional classes was thirty-nine percent. Teachers in argument-based inquiry classes spent less class time lecturing or guiding students. The total average talk time across the teachers in argument-based inquiry classes was eleven point forty-nine percent. Overall, the teachers' talk time in argument-based inquiry classes decreased over time from fourteen point seventeen percent in the first video clips to nine point fifty-two percent in the final video clips. In the first semester, the teacher talk time was fourteen point seventeen percent. In the second semester, the teachers' talk time in argument-based inquiry classes decreased to eleven point fifty-three percent. In the following two semesters (third and fourth), the teacher talk time decreased to ten point sixty-one percent and nine point fifty-two percent respectively. The teachers' talk time in traditional classes fluctuated across the duration of the project. In the first semester, the teachers talked forty-two point eight percent during class. In the second semester, the teachers talked more than the first semester (forty-three point thirty-six percent). In the third semester, the teachers' talk time decreased to thirty-six point six zero percent. In the fourth semester, the talk time increased to thirty-nine percent.

Summary of the results of teacher talk time

There was a clear difference in teacher talk time between the treatment and control groups across the duration of the period and at each time point. The treatment group teachers talked less across the duration of the project. In other words, the overall talk time across the duration of the period was shorter for the treatment group teachers than the control group teachers.

Figure 4. 1. Percentage of teacher talk time between the treatment and control group across semesters



In detail, the treatment group teacher talk time was shorter at each time point than the control group teachers. To sum up, the treatment group teachers consistently talked less than the control group teachers across the duration of the project and at each time point.

Structure of questions

Table 4. 2. shows the total percentages of question types (open-ended and close-ended) asked during class in both traditional and argument-based inquiry classes.

Table 4. 2. Average percentage of question types asked in traditional and argument-based inquiry classes

Group Type	Traditional class	Argument-based inquiry class
Open-ended question	22 %	56%
Close-ended question	78 %	44%

A clear difference emerged between traditional and argument-based inquiry classes. Traditional classroom teachers used close-ended questions most of the time compared to argument-based inquiry classroom teachers. Argument-based inquiry teachers used open-ended questions more frequently than close-ended questions. In traditional classes, teachers asked close-ended questions more than open-ended questions by a ratio of seventy-eight percent to twenty-two percent. In argument-based inquiry classes, the open-ended questions accounted for fifty-six percent and the close-ended questions accounted for forty-four percent.

A detailed analysis for the structure of questions is presented in tables 4. 3, 4. 4, and 4. 5. Table 4. 3 shows the percentage of open-ended questions and close-ended questions across classes in traditional and argument-based inquiry classes throughout the duration of the project (from the first semester to the forth semester). This level of analysis indicated that the pattern of the occurrence between open-ended questions and close-ended questions across classes during the project substantiated the consistence of teacher questioning differences.

Table 4. 3. Average percentage of question types asked in traditional and argument-based inquiry classes across the duration of the project

Group Type (semester)	Traditional class	Argument-based inquiry class
First year		
Open-ended question	5 %	23%
Open-ended question	5 %	12%
Second year		
Open-ended question	6%	10%
Open-ended question	6%	11%
First year		
Close-ended question	26%	17%
Close-ended question	17%	9%
Second year		
Close-ended question	21%	9%
Close-ended question	14%	9%
Total percentage of question types in each group	100%	100%

A difference between traditional and argument-based inquiry classes was that open-ended questions occurred more frequently and the percentage of open-ended questions was higher than the percentage of close-ended questions throughout the semesters in the argument-based inquiry classes. In traditional classes, the percentage of close-ended questions and open-ended questions varied throughout the semesters. The percentage of close-ended questions was twenty-six percent during the first semester, decreased to seventeen percent in the second semester, increased to twenty-one percent in the third semester, and decreased to fourteen percent in the fourth semester. The percentage of open-ended questions showed a different pattern than the percentage of close-ended question. The percentage of open-ended questions stayed at five percent throughout the first two semesters and increased to six percent in the third semester and the fourth semester.

In argument-based inquiry classes, the percentage of open-ended questions was twenty-three percent during the first semester, twelve percent during the second semester, ten percent during the third semester and eleven percent during the fourth semester. The percentage of close-ended questions was the highest in the first semester at seventeen percent but the percentage of close-ended questions decreased to nine percent across the rest of the three semesters.

Table 4. 4 shows the proportion of open-ended questions versus close-ended questions during each semester in traditional and argument-based inquiry classes throughout the duration of the project. The difference between table 4. 3 and table 4. 4 is that the percentages in table 4. 3 represents the percentage of open-ended questions and close-ended questions across semesters in the traditional group throughout the duration of the project and the percentage of open-ended questions and close-ended questions across semesters in the argument-based inquiry group throughout the duration of the project. The percentages in each column (traditional or argument-based inquiry group) add up to one hundred percent and the percentages in table 4. 4 represent the percentages of open-ended and close-ended questions as a pair in each semester in the traditional group or the argument-based inquiry group. The percentage of open-ended and close-ended questions in each semester in each group add up to one hundred percent. This level of analysis demonstrated a consistent pattern of the proportion of open-ended questions and close-ended questions that occurred in each group during each semester compared to table 4. 3 that accounted for the percentage of open-ended questions and closed-ended-questions in each group during the period of the project rather than during each semester.

Table 4. 4. Average percentage of question types asked in traditional and argument-based inquiry classes in each semester

Group Type (semester)	Traditional class	Argument-based inquiry class
First year		
Open-ended question (fall)	15 %	58%
Close-ended question (fall)	85 %	42%
Open-ended question (spring)	22%	57%
Close-ended question(spring)	78%	43%
Second year		
Open-ended question (fall)	23%	55%
Close-ended question (fall)	77%	45%
Open-ended question (spring)	28%	55%
Close-ended question (spring)	72%	45%

A consistent difference emerged between traditional and argument-based inquiry classes. Traditional classroom teachers asked more close-ended questions than open-ended questions in each semester compared to argument-based inquiry classroom teachers. The percentage of asking close-ended questions was eighty-five percent in the first semester, seventy-eight percent in the second semester, seventy-seven percent in the third semester and seventy-two percent in the fourth semester. In contrast, argument-based inquiry classroom teachers asked open-ended questions over fifty percent during each semester. The percentage of asking pen-ended questions was fifty-eight percent in the first semester, fifty-seven percent in the second semester, fifty-five percent in the third semester and fifty-five percent in the fourth semester.

Table 4. 5 shows the percentage of different sub-types of open-ended questions and close-ended questions in traditional and argument-based inquiry classes throughout the duration of the project. This level of analysis demonstrated the details of open-ended questions and close-ended questions across classes in both groups of teachers. Open-ended questions were comprised of three sub-types: (1) asking for explanation (AE), (2) asking for self-evaluation of reasoning

(AF), and (3) asking for self-evaluation of other's reasoning (AFO). Close-ended questions had two sub-types: (1) asking for factual information (AI), and (2) asking for confirmation (AC).

Table 4. 5. Average percentage of open-ended question and close-ended question types in traditional and argument-based inquiry classes across the duration of the project

Semester / Type	Group		Traditional class	Argument-based inquiry class
First	Open-ended question	AE	73%	46%
		AF	13%	31%
		AFO	14%	23%
	Close-ended question	AI	87%	72%
		AC	13%	28%
Second	Open-ended question	AE	67%	40%
		AF	21%	26%
		AFO	12%	34%
	Close-ended question	AI	89%	72%
		AC	11%	28%
Third	Open-ended question	AE	74%	52%
		AF	16%	22%
		AFO	10%	26%
	Close-ended question	AI	85%	74%
		AC	15%	26%
Forth	Open-ended question	AE	59%	56%
		AF	24%	17%
		AFO	17%	27%
	Close-ended question	AI	74%	72%
		AC	26%	28%

AE: asking for explanation, **AF**: asking for self-evaluation of reasoning, **AFO**: asking for self-evaluation of others' reasoning, **AI**: asking for information, and **AC**: asking for confirmation

There were two primary differences between traditional and argument-based inquiry classes in terms of open-ended questions and close-ended questions.

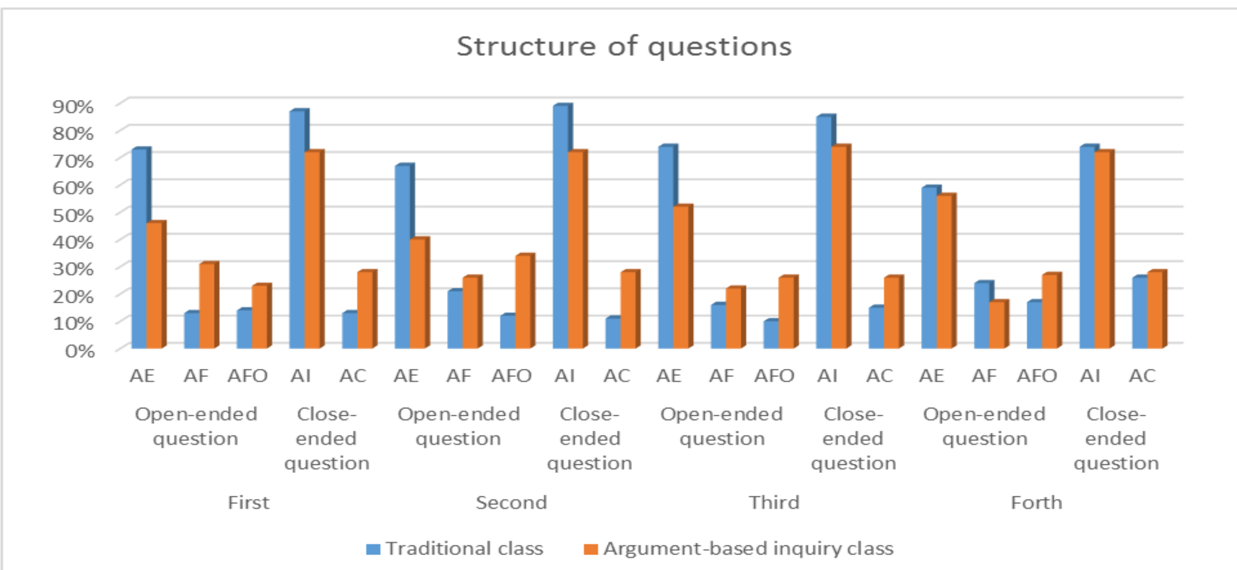
The first difference between traditional and argument-based inquiry classes was that the frequency of the occurrence of each open-ended sub-type question. Traditional classroom teachers used more asking for explanation (AE) questions than argument-based inquiry classroom teachers across the duration of the project. Asking for explanation was seventy-three percent in the first semester, sixty-seven percent at the second semester, seventy-four percent in the third semester, and fifty-nine percent in the fourth semester. Asking for explanation in argument-based inquiry classes was forty-six percent in the first semester, forty percent in the second semester, fifty-two percent in the third semester, and fifty-six percent in the fourth semester. In contrast, argument-based inquiry teachers used more asking for self-evaluation of other's reasoning (AFO) questions compared to traditional classroom teachers across semesters. Asking for self-evaluation of other's reasoning in argument-based inquiry classes was twenty-three percent in the first semester, thirty-four percent in the second semester, twenty-six percent in the third semester and twenty-seven percent in the fourth semester. Asking for self-evaluation of other's reasoning in traditional classes was fourteen percent in the first semester, twelve percent in the second semester, ten percent in the third semester, and seventeen percent in the fourth semester. Asking for self-evaluation of reasoning (AF) questions occurred more frequently in argument-based classes across the duration of the project except for the fourth semester. Traditional classroom teachers used asking for self-evaluation of reasoning seven percent higher than argument-based inquiry classroom teachers in the fourth semester. In the fourth semester, asking for self-evaluation of reasoning was used more frequently in traditional classes than argument-based inquiry classes. The occurrence of asking for self-evaluation in argument-based inquiry classes was thirty-one percent in the first semester, twenty-six percent in the second semester, twenty-two percent in the third semester, and seventeen percent in the fourth semester. The percentage of asking for self-evaluation in traditional classes was thirteen percent in the first semester, twenty-one percent in the second semester, sixteen percent in the third semester, and twenty-four percent in the fourth semester.

The second difference was the frequency of each sub-type of close-ended question. Traditional classroom teachers asked for factual information (AI) more frequently than argument-based inquiry classroom teachers across the duration of the project. The percentage of asking for factual information in traditional classes was eighty-seven percent in the first semester, eighty-nine percent in the second semester, eighty-five percent in the third semester, and seventy-four percent in the fourth semester. The percentage of asking for factual information in argument-based inquiry classes was seventy-two percent in the first semester, seventy-two percent in the second semester, seventy-four percent in the third semester, and seventy-two percent in the fourth semester. In contrast, argument-based inquiry classroom teachers used more asking for confirmation than traditional classroom teachers across semesters. Asking for confirmation in argument-based inquiry classes was twenty-eight percent in the first semester, twenty-eight percent in the second semester, twenty-six percent in the third semester, and twenty-eight percent in the fourth semester. Asking for confirmation in traditional classes was thirteen percent in the first semester, eleven percent in the second semester, fifteen percent in the third semester, and twenty-six percent in the fourth semester.

Summary of the results of structure of questions

A clear difference in the structure of questions (open-ended and close-ended questions) between traditional and argument-inquiry based groups emerged based on the frequency and percentage of each question type between the two groups, between the two groups at each time point and within each group at each time point in both as descriptive analyses and interpretive analyses (repeated measures ANOVA and effect sizes). The descriptive analyses (frequencies and percentages) indicated that there was a difference in the structure of questions between the two groups in terms of the differences in the frequencies and percentages of question types between the two groups. These results led to interpretive analyses to find if that difference was statistically significant. Overall, the frequency of open-ended questions was higher in the treatment group than in the control group across the duration of the project and at each time point. The treatment group teachers asked more open-ended questions than the control group teachers overall and at each time point. The control group teachers asked more close-ended questions overall and at each time point than the treatment group teachers. The frequency of question types was converted into percentages to control for the different duration of class times and to evaluate the patterns between the two groups in terms of the structure of questions (question types).

Figure 4. 2. Percentage of structure of questions between the treatment and control group across semesters



This figure shows trends for sub-question types in open-ended and close-ended questions between the two groups. With respect to sub open-ended question types, the treatment group teachers used asking for explanation (AE), asking for self-evaluation of reasoning (AF) and asking for self-evaluation of others' reasoning (AFO) more frequently that the control group teachers overall and at each time point except for time point 3 (third semester). At time point 3, asking for explanation (AE) was used more frequently by the control group teachers.

Student responses

Table 4. 6 shows the total average percentage of lower order and higher order student responses in traditional and argument-based inquiry classes.

Table 4. 6. Average percentage of student responses in traditional and argument-based inquiry classes

Group Type	Traditional class	Argument based inquiry class
Lower order thinking	56%	47%
Higher order thinking	44%	53%

A clear difference emerged between traditional and argument-based inquiry classes. Traditional classroom students responded using lower order thinking more often than argument-based inquiry classroom students. They used lower order thinking responses fifty-six percent of the time and higher order thinking responses forty-four percent of the time across the duration of the project. In contrast, argument-based inquiry classroom students used higher order thinking responses fifty-three percent of the time and lower order thinking responses forty-seven percent of the time.

Detailed analysis of student responses is presented in tables 4. 7, 4. 8, and 4. 9. Table 4. 7 shows the percentage of student responses across classes in traditional and argument-based inquiry classes across semesters or during each semester of the project. The difference between table 4. 7 and table 4. 8 is that the percentages in table 4. 7 represent the percentage of student responses across semesters in each group and the percentage in table 4. 8 represent the percentage of student responses in each semester in each group.

Table 4. 7. Average percentage of student responses in traditional and argument-based inquiry classes during each semester of the project

Group Type (semester)	Traditional class	Argument based inquiry class
Lower order thinking (first)	28%	17%
Lower order thinking (second)	18%	8%
Lower order thinking (third)	20%	8%
Lower order thinking (fourth)	15%	8%
Higher order thinking (first)	4%	22%
Higher order thinking (second)	4%	13%
Higher order thinking (third)	6%	11%
Higher order thinking (forth)	5%	13%
Total percentage of student responses in each group	100%	100%

A consistent difference emerged between traditional and argument-based inquiry classes. Traditional classroom students responded more using lower-order thinking than higher order thinking in each semester compared to argument-based inquiry classroom students. The relative percentage ratio of responses using lower order and higher order thinking at each time point was twenty-eight to four percent in the first semester, eighteen to four percent in the second semester, twenty to six percent in the third semester and fifteen to five percent in the fourth semester. In contrast, argument-based inquiry classroom students responded using higher-order thinking over at least ten percent during each semester. The relative percentage ratio of responses using lower and higher-order thinking was seventeen to twenty-two percent in the first semester, eight to thirteen percent in the second semester, eight to eleven percent in the third semester and eight to thirteen percent in the fourth semester.

Table 4. 8 shows the relative percentage of lower order and higher-order responses at each semester in traditional and argument-based inquiry classes.

Table 4. 8. Average percentage of student responses in traditional and argument-based classes during each semester

Group Type (semester)	Traditional class	Argument based inquiry class
Lower order thinking (first)	88%	43%
Higher order thinking (first)	12%	57%
Lower order thinking (second)	81%	40%
Higher order thinking (second)	19%	60%
Lower order thinking (third)	79%	42%
Higher order thinking (third)	21%	58%
Lower order thinking (forth)	74%	39%
Higher order thinking (forth)	26%	61%

A difference emerged between traditional and argument-based inquiry classes. Traditional classroom students more responded more frequently using lower order thinking than higher order thinking in each semester compared to argument-based inquiry classroom students. The percentage of lower order thinking responses was eighty-eight in the first semester, eighty-one percent in the second semester, seventy-nine percent in the third semester and seventy-four percent in the fourth semester. In contrast, argument-based inquiry classroom students responded using higher order thinking over fifty percent during each semester. The percentage of higher order thinking responses was fifty-seven percent in the first semester, sixty percent in the second semester, fifty-eight percent in the third semester and sixty-one percent in the fourth semester.

Table 4. 9 shows an average percentage of sub types of lower order and higher order thinking responses. This level of analysis demonstrated the details of lower order and higher order thinking responses across classes in both groups of teachers. Lower order thinking responses were comprised of one sub-type: (1) simple responses (S). Higher order thinking responses had three sub-types: (1) explanation responses (E), (2) self-evaluation of reasoning responses (SE), and (3) self-evaluation of others' reasoning responses (SEO).

Table 4. 9. Average percentage of higher order thinking response types in traditional and argument-based inquiry classes across the duration of the project

Semester / Type	Group		Traditional class	Argument-based inquiry class
First	Higher order thinking	E	77%	53%
		SE	11%	28%
		SEO	12%	19%
Second	Higher order thinking	E	71%	47%
		SE	20%	25%
		SEO	9%	28%
Third	Higher order thinking	E	80%	55%
		SE	12%	21%
		SEO	8%	24%
Fourth	Higher order thinking	E	62%	61%
		SE	22%	16%
		SEO	16%	23%

E: explanation response, **SE**: self-evaluation of reasoning response, and **SEO**: self-evaluation of others' reasoning response

There was one primary difference between traditional and argument-based inquiry classes in terms of lower order thinking and higher order thinking responses. Because lower order thinking responses were comprised of one sub type, there was no reason to differentiate its usage in terms of sub types. Referring to tables 4. 6, 4. 7 and 4. 8 demonstrates how many lower order thinking responses were used by students in both groups related to how many higher order thinking responses were used.

The first difference between traditional and argument-based inquiry classes was that the frequency of the occurrence of higher order thinking responses. Traditional classroom students used more explanation responses (E) than argument-based inquiry classroom students across the duration of the project. Explanation responses were seventy-seven percent in the first semester, seventy-one percent in the second semester, eighty percent in the third semester, and sixty-two percent in the fourth semester. Explanation responses (E) in argument-based inquiry classes were twenty-eight percent in the first semester, twenty-five percent in the second semester, twenty-one percent in the third semester, and sixteen percent in the fourth semester. In contrast, argument-based inquiry students used more self-evaluation of others' reasoning responses (SEO) compared to traditional classroom students across semesters. Self-evaluation of others' reasoning responses (SEO) in argument-based inquiry classes was nineteen percent in the first semester, twenty-eight percent in the second semester, twenty-four percent in the third semester and twenty-three percent in the fourth semester. Self-evaluation of others' reasoning responses in traditional classes was twelve percent in the first semester, nine percent in the second semester, eight percent in the third semester, and sixteen percent in the fourth semester. Self-evaluation of reasoning responses (SE) occurred more frequently in argument-based classes across the duration of the project except for the fourth semester. Traditional classroom students used self-evaluation of reasoning six percent higher than argument-based inquiry classroom students in the fourth semester. In the fourth semester, self-evaluation of reasoning responses (SE) were used more frequently in traditional classes than argument-based inquiry classes. The occurrence of self-evaluation responses in argument-based inquiry classes was twenty-eight percent in the first semester, twenty-five percent in the second semester, twenty-one percent in the third semester, and sixteen percent in the fourth semester. The percentage of self-evaluation of reasoning responses in traditional classes was eleven percent in the first semester, twenty percent in the second semester, twelve percent in the third semester, and twenty-two percent in the fourth semester.

Summary of the results of student responses

There was a clear difference in student responses (higher-order thinking and lower-order thinking) between the treatment and control group students across the duration of the project and at each time point. The frequency of higher-order thinking in the treatment group students was higher than in the control group students overall and at each time point. The treatment group students responded more frequently with higher-order thinking than the control group students overall and at each time point.

Figure 4. 3. Percentage of lower-order and higher-order student responses between the treatment and control group across semesters

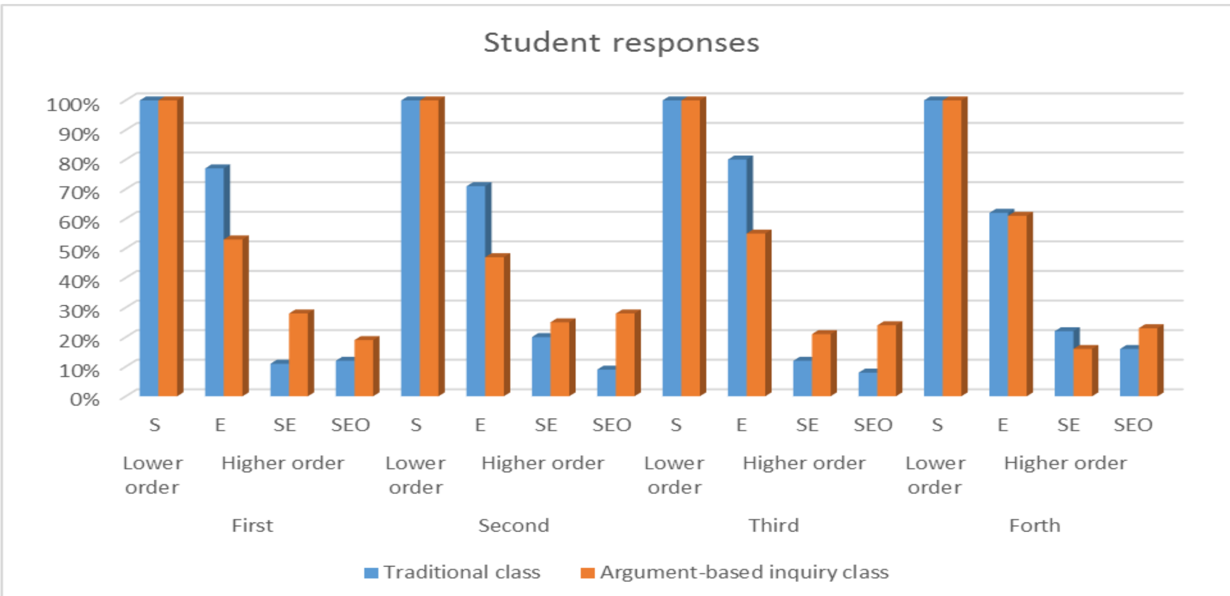


Figure 4. 3 shows patterns for sub-types of student responses between the two groups. With respect to higher-order thinking sub types, the treatment group students responded with self-evaluation of reasoning responses (SE), and self-evaluation of others' reasoning responses (SEO) more frequently than the control group students across the duration of the project and at each time point. However, the frequency of response explanation responses (E) in the control group students was higher than in the treatment group students overall and at each time point.

There was only one lower-order thinking type (simple response). There was no comparison within the lower-order thinking category. The frequency of simple responses in the control group students was higher than in the treatment group students overall and at each time point.

Relations between teacher question types and student response types

In order to investigate the relations between the frequency of question types and the frequency of student response types at each time point in detail, the ratios between the frequency of question types and the frequency of student response types in the treatment and control groups at each time point were calculated in percentages by providing row data (i.e., the frequency of each question type and student response types in both groups at each time point). Table 4. 10 and 4. 11. illustrate the details of the relations between question types and student response types.

Table 4. 10. The ratios between the frequency of open-ended questions and the frequency of higher-order thinking responses in the treatment and control groups at each time point

Group Semester	Treatment group	Control group
First year (Fall)	562/672 (84%)	110/160 (69%)
First year (Spring)	312/350 (89%)	124/162 (77%)
Second year (Fall)	286/302 (95%)	161/219 (74%)
Second year (Spring)	313/331 (95%)	151/189 (80%)

The ratios in parentheses were calculated in percentages by dividing the frequency of higher-order thinking responses (numerators) by the frequency of open-ended questions (denominators)

Table 4. 10. shows that the relations between the frequency of open-ended questions and the frequency of student higher-order thinking responses in the treatment and control groups at each time point. The more frequently open-ended questions were asked, the more frequently student higher-order thinking responses were generated accordingly in the treatment group compared to the control group. The frequency of open-ended-questions and the ratios of the frequency of open-ended questions to the frequency of student higher-order student responses were directly proportional.

Table 4. 11. The ratios between the frequency of close-ended questions and the frequency of lower-order thinking responses in the treatment and control groups at each time point

Group / Semester	Treatment group	Control group
First year (Fall)	423/486 **(87)**	801/874 **(92)**
First year (Spring)	209/259 **(81)**	513/585 **(88)**
Second year (Fall)	203/249 **(82)**	591/722 **(82)**
Second year (Spring)	201/273 **(74)**	430/487 **(88)**

The ratios in parentheses were calculated in percentages by dividing the frequency of lower-order thinking responses (numerators) by the frequency of close-ended questions (denominators)

Table 4. 11. shows that the relations between the frequency of close-ended questions and the frequency of student lower-order thinking responses in the treatment and control groups at each time point. The more frequently close-ended questions were asked, the more frequently student lower-order thinking responses were generated accordingly in the control group compared to the treatment group. The frequency of close-ended-questions and the ratios of the frequency of close-ended questions to the frequency of student lower-order student responses were directly proportional.

Summary of the descriptive analysis results

The treatment group teachers talked less than the control group teachers overall and at four different time points. The treatment group teachers asked open-ended questions more frequently than the control group teachers. In addition, the control group teachers asked close-ended questions more frequently than the treatment group teachers. The treatment group students responded by more frequently using higher-order thinking than the control group students.

Interpretive analysis results

Based on the descriptive analysis results, interpretive statistical analyses were conducted. The first step was to conduct repeated measures ANOVA for the three categories: (1) teacher talk time, (2) question types, and (3) student responses. In order to interpret the results of the statistical analyses accurately, due to the violation of the assumption of Sphericity, statistical tests with corrected degrees of freedom were used.

In the second step, the relation between teacher talk time and the two categories (structure of questions and student responses) that emerged during the coding of the teachers' video clips was identified based on the results of repeated measures ANOVA and the effect sizes for the three categories between the two groups at each time point.

The third step was to identify the relation between talk time and each of the two categories (structure of questions and student responses) based on the results of repeated measures ANOVA and the effect sizes for the three categories between the two groups at each time point.

The last step was to identify the relation between the two categories (structure of questions and student responses) based on the results of repeated measures ANOVA and the effect sizes for the two categories between the two groups at each time point.

Teacher talk time

Talk time was analyzed using repeated measures ANOVA to find the difference between the treatment and control groups over the period of the project and at each time point, and the difference within each group at each time point. The repeated measures ANOVA were conducted at the alpha level = .05. Table 1 in appendix B shows the assumption of Sphericity for teacher talk time using Maulchly's test. The assumption of Sphericity should be met to obtain valid results, in terms of accurate F values, from repeated measures ANOVA.

The test result indicated that the assumption has been violated because the significance figure for teacher talk time was $\chi^2 = 12.026$, $p < .034$. In other words, there was a difference between the variances of differences between all of the combinations of experimental conditions (four different time points). To verify the results for this test, corrected degrees of freedom were calculated so that the accurate F value was obtained. Table 2 in appendix B shows the results of tests with corrected degrees of freedom for the violation of Sphericity.

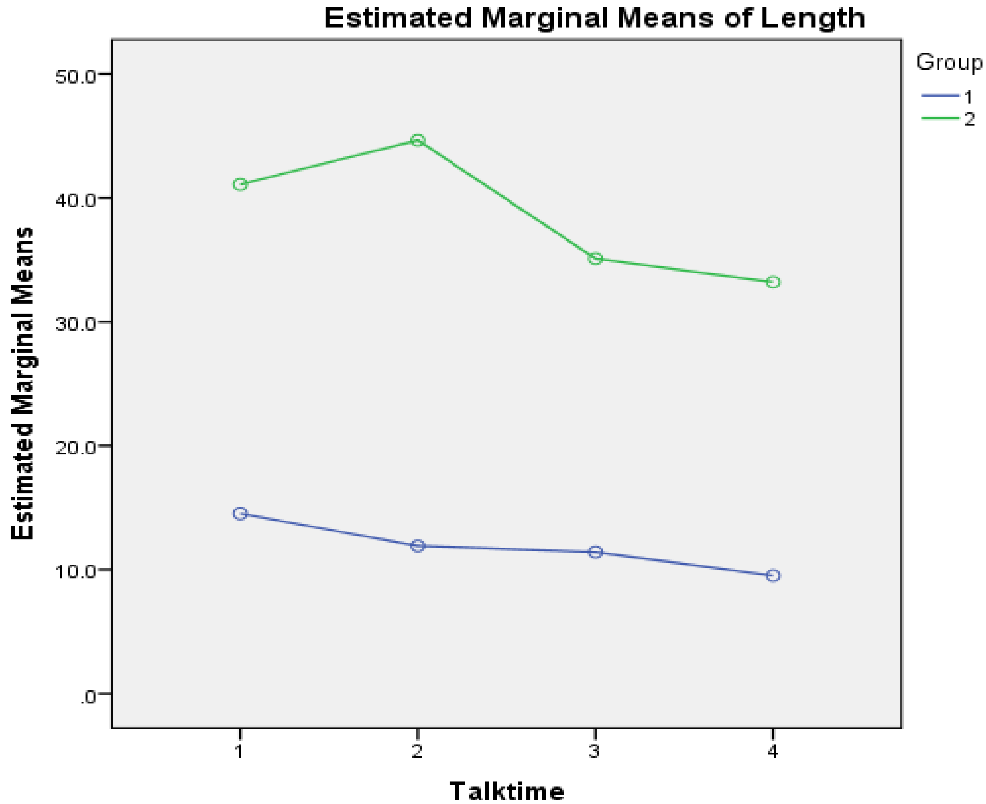
In table 1 in appendix B, Epsilon is greater than 0.75 in Greenhouse-Geisser so Huynh-Feldt needs to be used in table 2 in appendix B (Howell, 1992). Based on the corrected degrees of freedom in order to have the F-value accurate, there was a significant difference in teacher talk time for both the treatment and control groups at four different time points because the significant value was less than 0.05, which was $F(2.87, 146.12) = 6.61$, $p < .001$. There was no evidence of an interaction between group and talk time at each time point because the significant value was .058, which was $F(2.87, 146.12) = 2.59$, $p > 0.58$. It indicated that teacher talk time in both groups at each time point did not change in different directions. The trends for the both groups were similar at each time point. However the value was close to the alpha level (the normal criterion of .050) so if the sample size was bigger, there is a chance that there would be evidence that there is an interaction between time and group.

Table 3 in appendix B shows the difference between groups (treatment and control groups). There was an overall significant difference in teacher talk time between the two groups, $F(1, 51) = 445.59$, $p < .001$.

Summary of teacher talk time using repeated measures ANOVA

There was a significant difference in teacher talk time between the treatment and control groups at each time point as well as an overall difference between the treatment and control groups. Graph 4. 1 shows the trends in teacher talk time for both groups at each time point.

Graph 4. 1. Trends in teacher talk time for the treatment and control groups



The graph shows patterns for both groups at each time point. The treatment group teachers talked less than the control group teachers at different time points. The two groups changed over time. The overall difference between the two groups was significant. The graph doesn't indicate both groups' teacher talk time changed in different patterns at each time point. In other words, the patterns for both groups changed in a similar way.

Table 4. 12. Effect size between the treatment and control group teacher talk time at each time point

[Treatment group values – Control group values]

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	-2.81 (large)	-2.46 (large)	-1.81 (large)	-2.23 (large)

The effect sizes for teacher talk time between the two groups at each time point were calculated. The effect sizes for teacher talk time between the two groups at each time point started out at -2.81, increased to -1.81 until time point 3 and decreased at time point 4 to -2.23.

Structure of questions

Structure of questions was analyzed using repeated measures ANOVA to find the difference between the treatment and control groups over the period of the project and at each time point, and the difference within each group at each time point. The repeated measures ANOVA was conducted using the alpha level = .05. Table 17 shows the results of multivariate tests, also known as MANOVA.

Open-ended questions

There were two question types: open-ended questions and close-ended questions. The open-ended questions were analyzed first. Table 4 in appendix B shows the difference between the means of the two groups (treatment and control groups) on a combination of three measured dependent variables: (1) asking for explanation (E), (2) asking for self-evaluation of reasoning (AF) and (3) asking for self-evaluation of others' reasoning (AFO).

Among the four tests, Wilks' Lambda was used to find any significant difference between the means of the two groups and within each group on a combination of the three dependent variables because it is widely used for MANOVA tests (Everitt & Dunn, 1991; Polit, 1996). Table 4 in appendix B indicates three hypothesis test results for open-ended questions: (1) there was a significant difference between the means of the two groups on a combination of the three dependent variables, $F(3, 49) = 9.200$, $p<.01$, (2) there was a significant difference between the means within each group on a combination of the three dependent variables at each time point, F

(9, 43) = 2.800, $p<.011$ and (3) There was a significant difference in the interaction between time and group, F (9, 43) = 2.613, $p<0.17$. In other words, both groups changed in different directions over time.

Table 5 in appendix B indicates that the assumption of Sphericity for the three sub types of open-ended questions has been violated, which was asking for explanation (AE) with $\chi^2 = 27.957$, $p< .001$, asking for self-explanation of reasoning (AF) with $\chi^2 = 63.145$, $p< .001$, and asking for self-evaluation of others' reasoning (AFO) with $\chi^2 = 27.708$, $p< .001$. To obtain the accurate F values for the effects of the three sub open-ended question types, tests with corrected degrees of freedom for open-ended questions were conducted. Greenhouse-Geisser was less than .75 so Huynh-Feldt needed to be used (refer to table 6 in appendix B for the univariate tests)

Tests with corrected degrees of freedom for the violation of Sphericity shows six hypothesis test results: (1) there was a significant difference between the two groups on AE at each time point, (2) there was a significant difference between the two groups on AF at each time point (3) there was no significant difference between the two groups on AFO at each time point, (4) There was a significant difference on the interaction between time and group the two groups on AE at each time point. In other words, the trends for both groups on AE at each time point changed in different directions, (5) there was a significant difference on the interaction between time and group on AF at each time point. In other words, the trends for both groups on AF at each time point changed in different directions, and (6) there was no significant difference on the interaction between time and group on AFO at each time point. In other words, the trends for both groups on AFO at each time point changed in a similar pattern.

Table 7 in appendix B shows that there was a significant difference between the two groups in terms of AE, AF and AFO, which were three sub-open-ended question types.

Summary of open-ended questions

In order to evaluate the degree of the effect of the intervention for open-ended questions, effect sizes were calculated.

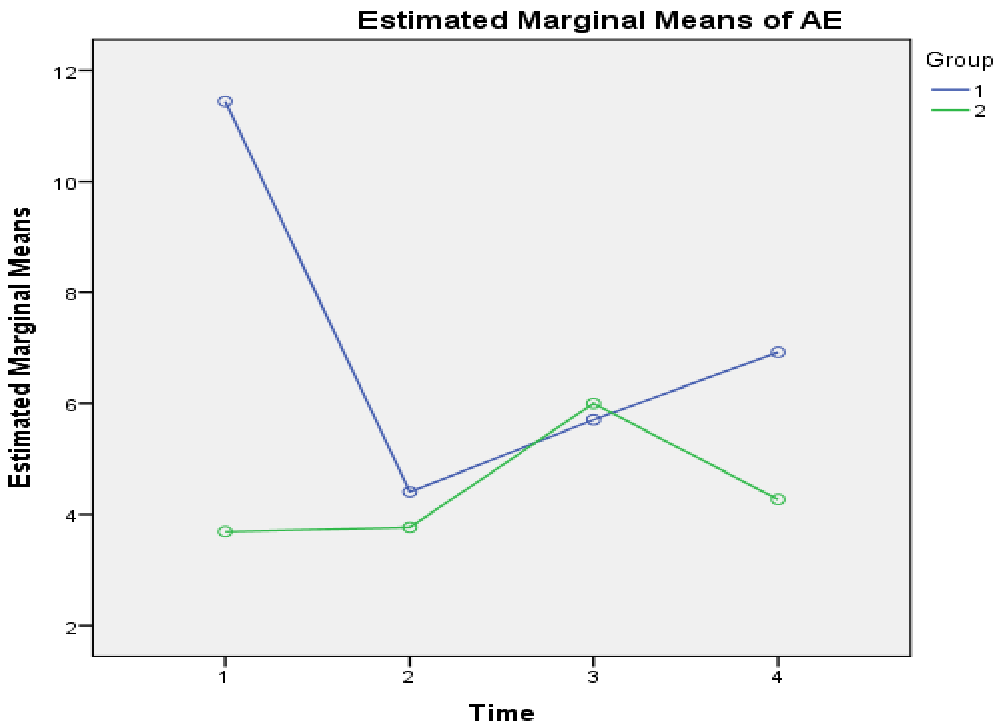
Table 4. 13. Effect size between the treatment and control group at each time point in open-ended question

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	1.04 (large)	0.76 (medium)	0.33 (small)	0.57 (medium)

The effect sizes for open-ended questions between the two groups at each time point started at 1.04, decreased to more than two-thirds of the starting effect size until time point 3 and increased back at time point 4 to 0.57. Overall, the effect sizes for open-ended questions between the two groups at different time points showed medium to large differences except for time point 3. Time point 3 showed a small degree of difference.

Summary of asking explanation (AE).

Graph 4. 2. Trends for asking for explanation (E) for the treatment and control groups



Graph 4. 2 shows that there was a significant difference between the two groups on asking for explanation (AE). However, the two groups showed different trends at each time point. The treatment group (group 1) teachers asked many AE questions at time point 1 (first semester), then the freaquency of AE decreased drastically at time point 2 (second semester), then the frequency of AE started to increase at time points 3 and 4 (third and fourth semesters). The control group (group 2) showed a different pattern. The frequency of AE was lowest at time point 1. The frequency started to increase at time points 2 and 3. At time point 3, the control group teachers asked more AE questions than the treatment group teachers. The frequency of AE started to decrease at time point 4.

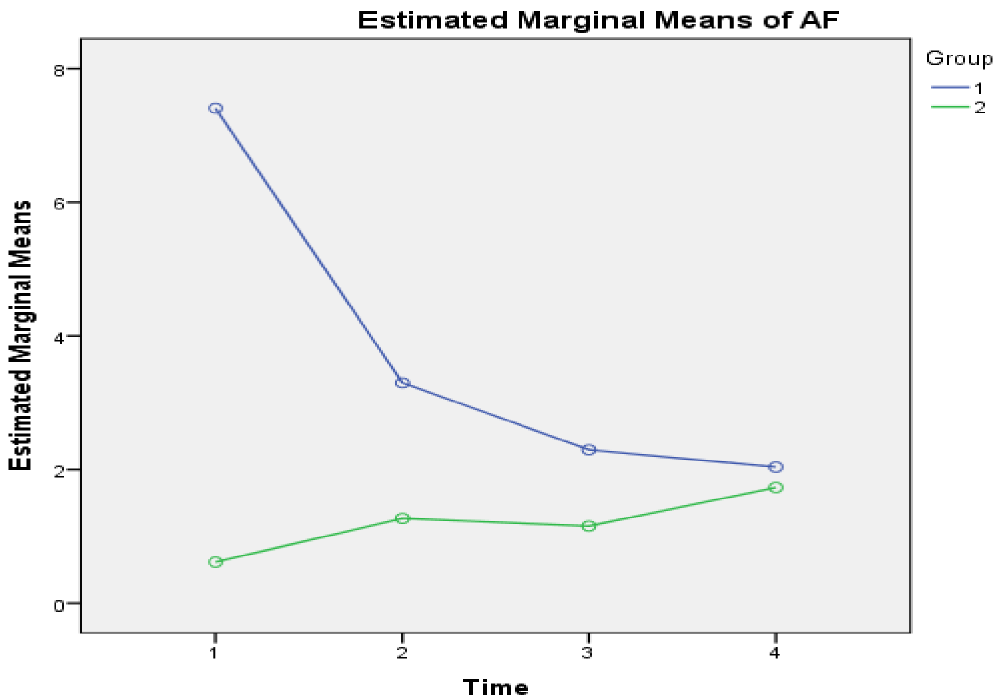
Table 4. 14. Effect size between the treatment and control group at each time point in asking for explanation (AE)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	0.80 (large)	0.25 (small)	-0.02 (small)	0.46 (small)

The effect sizes for asking for explanation (AE) between the two groups were calculated at different time points. The values for the treatment group at each time point were higher than the values for the control group except for time point 3. The effect sizes for asking for explanation (AE) between the two groups at each time point started out at 0.81, decreased to 0.02 until the sign of the effect size is reversed at time point 3 and rebounded back at time point 4 to 0.46. In order to explain the reverse of the values of the treatment and control groups at time point 3, the negative sign was added. Overall, the effect sizes for asking for explanation (AE) beween the two groups showed a small to large degree of difference except for time point 3.

Summay of asking for self-evaluation of reasoning (AF)

Graph 4. 3. Trends for asking for self-evaluation of reasoning (AF) for the treatment and control groups



Graph 4. 3 shows that there was a significant difference between the two groups on asking for self-evaluation of reasoning (AF). There were different trends for asking self-evaluation of reasoning between the two groups at each time point. The treatment group (group 1) showed a decreasing trend at each time point. The frequency of AF was highest at time point 1. The frequency of AF started to decrease considerably at time point 2 and continued to decrease gradually at time points 2 and 3. However, the control group (group 2) showed a different pattern. The frequency of AF was lowest at time point 1. The frequency of AF started to increase gradually at time point 2, then decreased a little at time point 3 and then started to increase gradually again at time point 4.

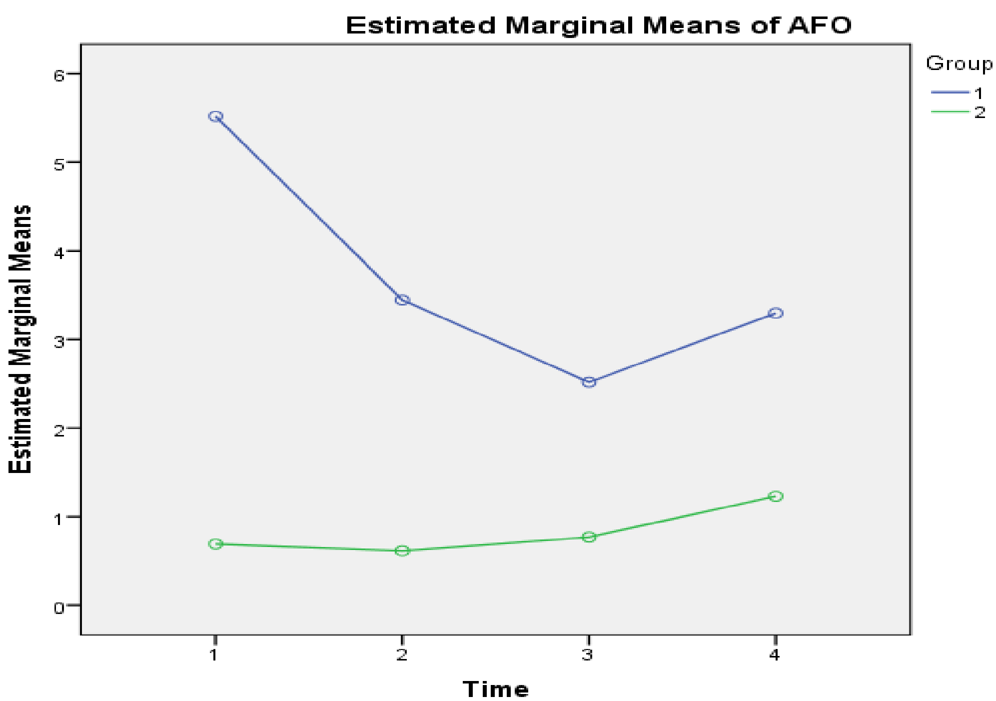
Table 4. 15. Effect size between the treatment and control group at each time point in asking for self-evaluation of reasoning (AF)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	1.16 (large)	0.62 (medium)	0.39 (small)	0.12 (small)

The effect sizes for asking for self-evaluation of reasoning (AF) between the two groups at each time point were calculated. The effect sizes for asking for self-evaluation of reasoning (AF) between the two groups at each time point started out at 1.16 and continued to decrease to 0.12 until time point 4. Overall, the effect sizes for self-evaluation of reasoning (AF) between the two groups at different time points showed a small to large degree of difference.

Summary of asking for self-evaluation of others' reasoning (AFO)

Graph 4. 4. Trends for asking for self-evaluation of others' reasoning (AFO) for the treatment and control groups



Graph 4. 4 shows there was a significant difference between the two groups on asking for self-evaluation of others' reasoning (AFO). There were different trends for asking evaluation of others' reasoning between the two groups at each time point. The treatment group (group 1) showed that the frequency of AFO was highest at time point 1, started to decrease considerably at time points 2 and 3, and then rebounded back significantly at time point 4. The control group (group 2) showed a different pattern. The frequency of AFO was lowest at time point 1 and then increased gradually at time points 2, 3 and 4.

Table 4. 16. Effect size between the treatment and control group at each time point in asking for self-evaluation of others' reasoning (AFO)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	0.84 (large)	0.86 (large)	0.60 (medium)	0.66 (medium)

The effect sizes for asking for self-evaluation of others' reasoning (AFO) between the two groups at each time point were calculated. The effect sizes for asking for self-evaluation of others' reasoning (AFO) between the two groups at each time point started out at 0.84, was 0.86 at time point 2, decreased to 0.60 at time point 3 and rebounded back slightly at time point 4 to 0.66. There was not much difference among the effect sizes at each time point. Overall, the effect sizes for asking for self-evaluation of others' reasoning (AFO) at different time points showed a medium to large degree of difference.

Close-ended questions

The second part of structure of questions was close-ended questions. Close-ended questions consisted of two sub-types: (1) asking for factual information (AI) and (2) asking for confirmation (AC).

Among the four tests, Wilks' Lambda was used to find any significant difference between the means of the two groups and within each group on a combination of the two dependent variables because it is widely used for MANOVA tests (Everitt & Dunn, 1991; Polit, 1996).

Table 8 in appendix B indicates three hypothesis test results for close-ended questions: (1) there was a significant difference between the means of the two groups on a combination of the two dependent variables, $F (2, 50) = 12.779, p<.01$, (2) there was a significant difference between the means within each group on a combination of the two dependent variables at each time point, $F (6, 46) = 2.472, p<.001$, and (3) There was a significant difference in the interaction between time and group at each time point, $F (6, 46) = 2.472, p<0.37$. In other words, there was a significant difference in the trends of both groups at each time point.

Table 9 in appendix B indicates that the assumption of Sphericity for AI has been violated, which was asking for factual information (AI) with $\chi^2 = 13.650, p< .018$, and asking for confirmation (AC) with $\chi^2 = 10.634, p< .059$. However, the assumption of Sphericity for AC has been met, $p>0.59$. To obtain the accurate F values for the effects of AI, tests with corrected degrees of freedom for AI questions were conducted. Greenhouse-Geisser is greater than .75 so Huynh-Feldt needs to be used (refer to table 10 in appendix B for the univariate tests)

Tests for AC and tests with corrected degrees of freedom for AI for the violation of Sphericity showed four hypothesis test results: (1) there was a significant difference between the two groups on AI at each time point, (2) there was a significant difference between the two groups on AC at each time point, (3) there was no significant in the interaction between time and group on AI at each time point. In other words, both groups changed in a similar pattern at each time point, and (4) there was no significant difference in the interaction between time and group on AC at each time point. In other words, both groups changed in a similar pattern at each time point.

Table 11 in appendix B shows that there was a significant difference between the two groups in terms of AI, which was F (1, 51) = 25.952, p<0.001 but there was no significant difference between the two groups on AC, which was F (1, 51) = 452, p> .51.

Summary of closed-ended questions

Effect sizes were calculated in order to evaluate the degree of the effect of the intervention for close-ended questions.

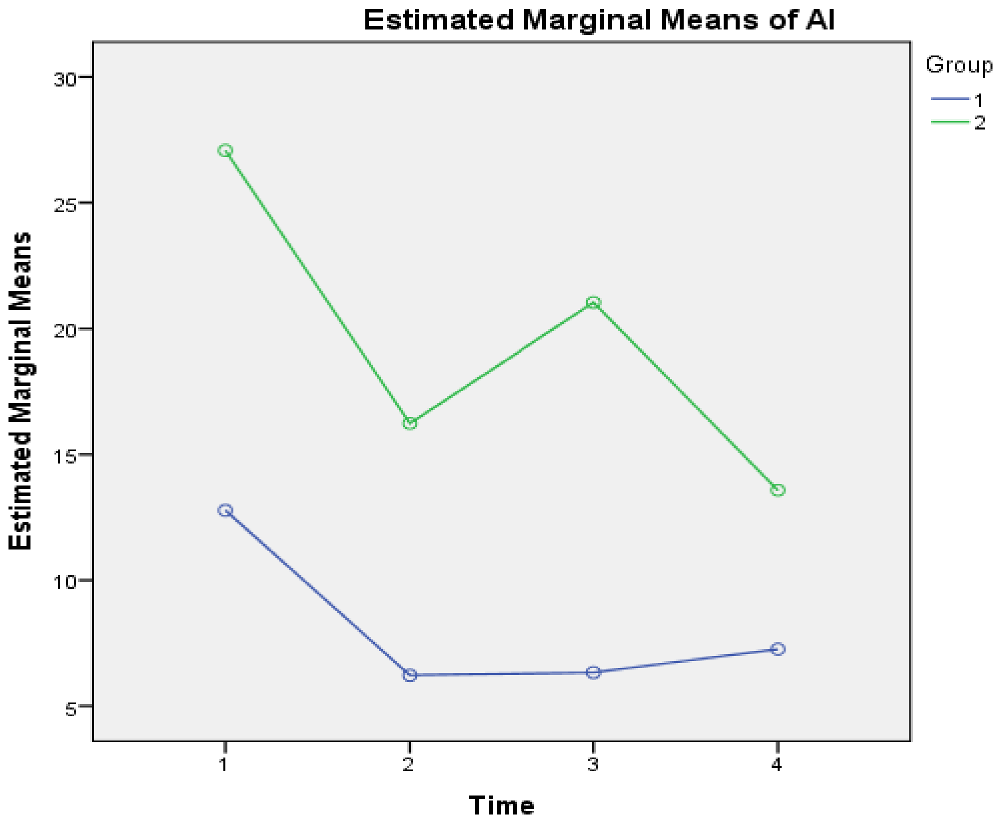
Table 4. 17. Effect size between the treatment and control group at each time point in close-ended questions

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	-0.75 (medium)	-0.93 (large)	-1.07 (large)	-0.74 (medium)

The effect sizes for closed-ended questions were calculated at different time points. The effect size for close-ended questions between the two groups at each time point started at -0.75, decreased to -1.07 until time point 3 and increased to -0.74 at time point 4. Overall, the effect sizes for close-ended questions between the two groups at each time point showed medium to large degrees of difference.

Summary of asking for factual information (AI)

Graph 4. 5. Trends for asking for factual information (AI) for the treatment and control groups



Graph 4. 5 shows that there was a significant difference between the two groups on asking for information (AI). However, there were different patterns on AI between the two groups at each time point. The treatment group (group 1) started out with the highest frequency of AI at time point 1, started to decrease considerably at time point 2, rebounded back at time point 3, and then decreased again at time point 4. The control group (group 2) started out relatively high with the highest frequency of AI at time point 1, decreased considerably at time point 2, then increased slightly at time points 3 and 4.

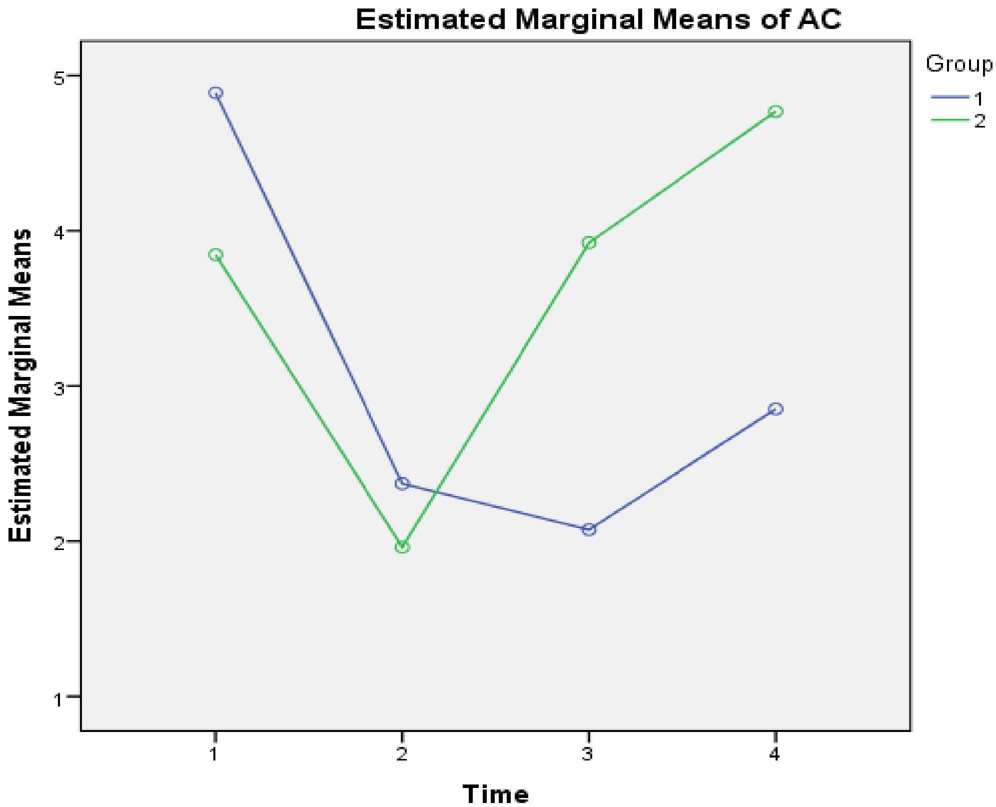
Table 4. 18. Effect size between the treatment and control group at each time point in asking for factual information (AI)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	-0.92 (large)	-2.25 (large)	-1.05 (large)	-0.71 (medium)

The effect sizes for asking for information (AI) between the two groups at each time point were calculated. The effect sizes for asking for information (AI) between the two groups at each time point started out at -0.92, decreased considerably to -2.25 at time point 2, increased by more than half to -1.05 at time point 3, and continued to increase to -0.71 at time point 4. Overall, the effect sizes for asking for information (AI) between the two groups at different time points showed medium to large degrees of difference.

Summary of asking for confirmation (AC)

Graph 4. 6. Trends for asking for confirmation (AC) for the treatment and control groups



Graph 4. 6 shows there was a significant difference between the two groups on asking for confirmation (AC). There were different patterns on AC between the two groups at each time point. The treatment group (group 1) had the highest frequency of AC at time point 1, decreased considerably at time point 2, continued to decrease slightly at time point 3, and then increased at time point 4. The control group (group 2) has the second highest frequency of AC at time point 1, then decreased conservatively at time point 2, rebounded back at time point 3 and continued to increase considerably at time point 4.

Table 4. 19. Effect size between the treatment and control group at each time point in asking for confirmation (AC)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	-1.64 (large)	0.11 (small)	-0.32 (small)	-0.35 (small)

The effect sizes for asking for confirmation (AC) between the two groups at each time point were calculated. At time point 2, the effect size was positive because the treatment group values were larger than the control group values. The effect sizes for asking for confirmation (AC) between the two groups at each time point started out low at -1.64, increased to 0.11 at time point 2, dropped back to -0.32 at time point 3 and decreased slightly to -0.35 at time point 4. Overall, the effect sizes for asking for confirmation (AC) between the two groups at different time points showed a small to large degree of difference regardless of the direction.

Summary of structure of questions

There were averages of medium or larger effects for open-ended and close-ended questions between the treatment and control groups at different time points. The treatment group teachers used open-ended questions more frequently than the control group teachers at each time point in terms of asking for explanation (AE), asking for self-evaluation of reasoning (AF) and asking for self-evaluation of others' reasoning (AFO). In contrast, the treatment group teachers used close-ended questions less frequently than the control group teachers in terms of asking for factual information (AI) and asking for confirmation (AC).

Student responses

Student responses were analyzed using repeated measures ANOVA to find the difference between the treatment and control groups over the period of the project and at each time point, and the difference within each group at each time point. The repeated measures ANOVA was conducted using alpha level = .05. Table 12 in appendix B shows the results of multivariate tests.

Higher-order thinking responses

There were two types of student responses: higher-order thinking and lower-order thinking. The higher-order thinking was analyzed first. The higher-order thinking consists of three sub-categories: (1) explanation responses (E), (2) self-evaluation of reasoning responses (SE), and (3) self-evaluation of others' reasoning responses (SEO).

Among the four tests, Wilks' Lambda was used to find any significant difference between the means of the two groups and within each group on a combination of the two dependent variables because it is widely used for MANOVA tests (Everitt & Dunn, 1991; Polit, 1996).

Table 12 in appendix B indicates three hypothesis test results for higher-order thinking responses: (1) there was a significant difference between the means of the two groups on a combination of the three dependent variables, $F\ (3, 49) = 10.166$, $p<.01$, (2) there was no significant difference between the means within each group on a combination of the two dependent variables at each time point, $F\ (9, 43) = 1.968$, $p>.067$, and (3) There was a significant difference in the interaction between time and group at each time point, $F\ (9, 43) = 2.421$, $p<0.25$. In other words, both group changed in different directions.

Table 13 in appendix B indicates that the assumption of Sphericity for E, SE, and SEO has been violated, which was explanation response (E) with $\chi^2 = 21.059$, $p< .001$, self-evaluation (SE) with $\chi^2 = 54.499$, $p< .001$, and self-evaluation of others' reasoning responses (SEO) with $\chi^2 = 14.316$, $p< .014$.To obtain the accurate F values for the effects of E, SE, and SEO, tests with corrected degrees of freedom were conducted. Greenhouse-Geisser was greater than .75 so Huynh-Feldt needed to be used (refer to table 14 in appendix B for the univariate tests).

Tests with corrected degrees of freedom for E, SE, and SEO for the violation of Sphericity shows six hypothesis test results: (1) there was a significant difference between the two groups on E at each time point, (2) there was a significant difference between the two groups on SE at each time point, (3) there was no significant difference between the two groups on SEO at each time point, (4) there was a significant difference in the interaction between time and group on E at each time point, (5) there was a significant difference in the interaction between time and group on SE at each time point, and (6) there was no significant difference in the interaction between time and group on SEO at each time point.

Table 15 in appendix B shows that there was a significant difference between the two groups in terms of E, SE, and SEO, which were $F(1, 51) = 14.35$, $p<.001$, $F(1, 51) = 16.00$, $p<.001$, $F(1, 51) = 18.56$, $p<.001$ respectively.

Summary of higher-order thinking responses

Effect sizes were calculated in order to evaluate the degree of the effect of the intervention for higher-order thinking responses.

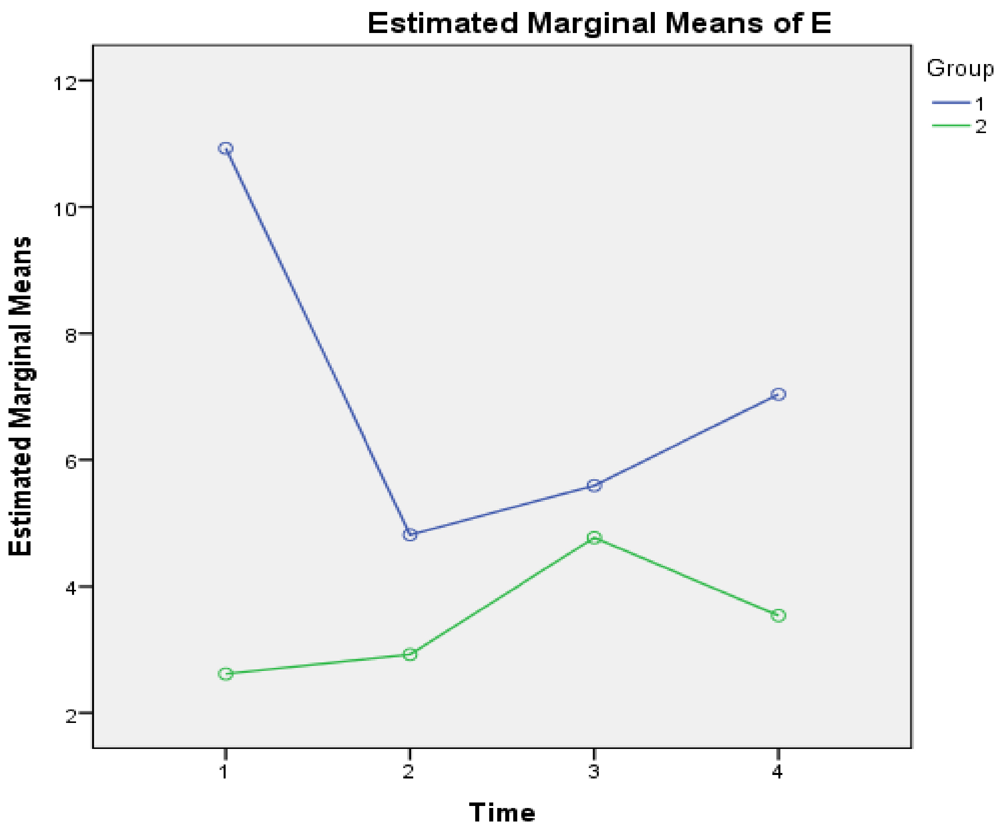
Table 4. 20. Effect size between the treatment and control group at each time point in higher-order thinking

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	1.14 (large)	0.86 (large)	0.59 (medium)	0.76 (medium)

The effect sizes for student higher-order thinking between the two groups at each time point were calculated. The effect sizes for student higher-order thinking between the two groups at each time point started out at 1.14, decreased by half until time point 3 and rebounded back at time point 4 to 0.76. Overall, the effect sizes for student higher-order thinking between the two groups at different time points showed a medium to large degree of difference.

Summary of explanation responses (E)

Graph 4. 7. Trends for Explanation responses (E) for the treatment and control groups



Graph 4. 7 shows that there was a significant difference between the two groups. However, there were different patterns of E between the two groups. The treatment group (group 1) had the highest frequency of E at time point 1, decreased considerably at time point 2, rebounded back at time point 3 and then increased at time point 4. The control group (group 2) had the lowest frequency of E at time point 1, increased gradually at time point 2, continued to

increase at time point 3 and then decreased at time point 4. The trends for E were a mirror of the trends for asking for explanation (AE) (refer to graph 4. 2).

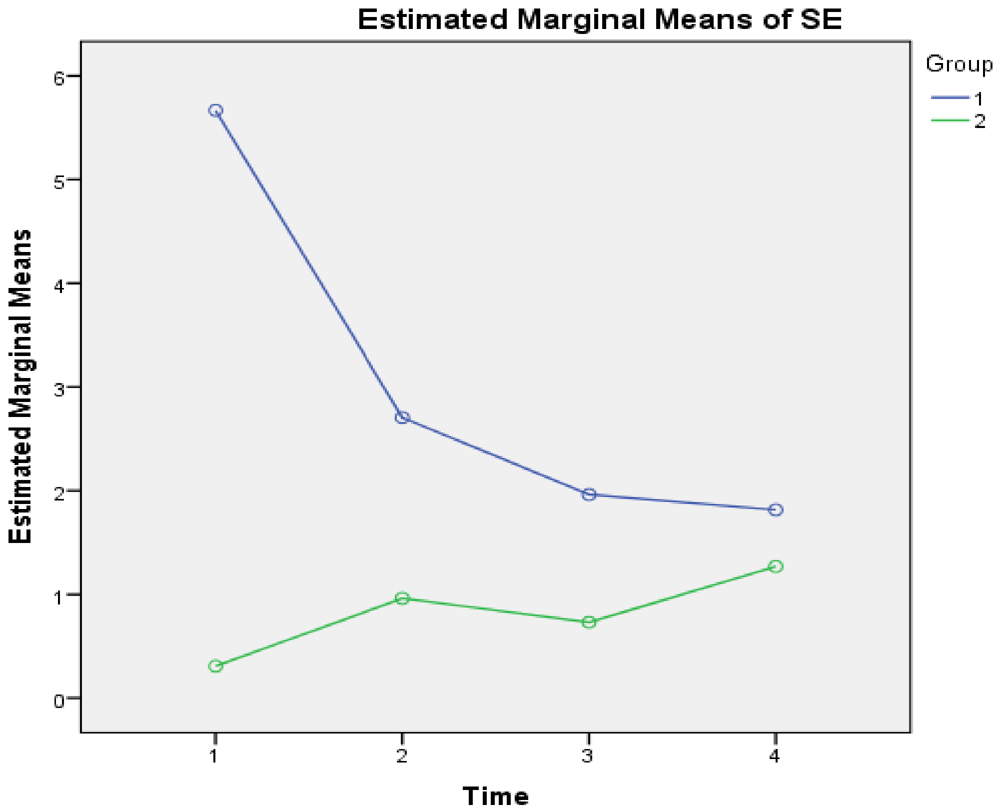
Table 4. 21. Effect size between the treatment and control group at each time point in explanation responses (E)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	0.95 (large)	0.51 (medium)	0.18 (small)	0.66 (medium)

The effect sizes for explanation responses (E) between the two groups at each time point were calculated. The effect sizes for explanation responses (E) between the two groups at each time point started out high at 0.95, continued to decrease to 0.18 until time point 3 and rebounded back at time point 4 to 0.66. The effect sizes for explanation response (E) between the two groups at different time points showed a small to large degree of difference.

Summary of self-evaluation of reasoning responses (SE)

Graph 4. 8. Trends for self-evaluation of reasoning (SE) for the treatment and control groups



Graph 4. 8 shows there was a significant difference of SE between the groups. However, there was a different trend between the two groups. The treatment group (group 1) had the highest frequency of SE at time point 1, decreased at time point 2, and then continued to decrease at time points 3 and 4. The control group (group 2) had the lowest frequency of SE at time point 1, increased at time point 2, decreased slightly at time point 3, and then rebounded again at time point 4. The trends for SE were a mirror of the trends for asking for self-evaluation of reasoning (AF) (refer to graph 4. 3).

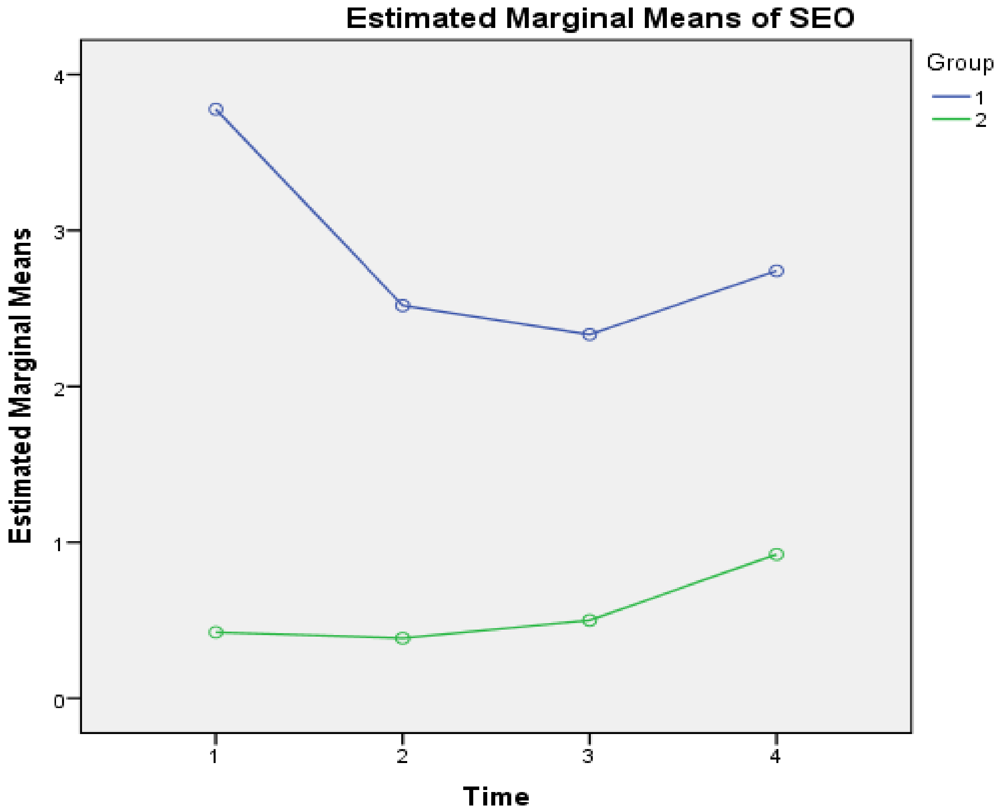
Table 4. 22. Effect size between the treatment and control group at each time point in self-evaluation of one's own reasoning (SE)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	1.09 (large)	0.56 (medium)	0.54 (medium)	0.25 (small)

The effect sizes for self-evaluation of one' own reasoning responses (SE) between the two groups at each time point were calculated. The effect sizes for self-evaluation of one' own idea responses (SE) between the two groups at each time point started out high at 1.09, continued to decrease by almost one fourth to 0.25 until time point 4. Overall, the effect sizes for self-evaluation of one's own reasoning (SE) showed a small to large degree of difference.

Summary of self-evaluation of others' reasoning (SEO)

Graph 4. 9. Trends for self-evaluation of others' reasoning (SEO) for the treatment and control groups



Graph 4. 9 shows there was a significant difference of SEO between the groups. However, there was a different pattern between the two groups. The treatment group (group 1) had the highest frequency of SEO at time point 1, decreased at time point 2, continued to decrease at time point 3, and then rebounded at time point 4. The treatment group (group 2) had the lowest frequency of SEO at time point 1, increased at time point 2, and then continued to increase at time points 3 and 4. The trends for SEO were a mirror of the trends for asking for self-evaluation of others' reasoning (AFO) (refer to graph 4. 4).

Table 4. 23. Effect size between the treatment and control group at each time point in self-evaluation of others' reasoning (SEO)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	0.85 (large)	0.79 (medium)	0.63 (medium)	0.60 (medium)

The effect sizes for self-evaluation of others' ideas responses (SEO) between the two groups at each time point were calculated. The effect sizes for self-evaluation of others' ideas responses (SEO) between the two groups at each time point started out at 0.85, and continued to decrease to 0.60 until time point 4. The overall pattern for the effect size in SEO was a decrease over the time. The effect sizes for self-evaluation of others' ideas responses (SEO) between the two groups at different time points showed a medium to large degree of difference.

Lower-order thinking responses

The second part of student responses was lower-thinking responses, which consists of simple responses (S).

Among the four tests, Wilks' Lambda was used to find any significant difference between the means of the two groups and within each group on a combination of the two dependent variables because it is widely used for MANOVA tests (Everitt & Dunn, 1991; Polit, 1996).

Table 16 in appendix B indicates three hypothesis test results for lower-order thinking responses: (1) there was a significant difference between the means of the two groups on one dependent variable at each time point, F (3, 49) = 6.808, $p<.01$, and (2) There was no significant difference in the interaction between time and group at each time point, F (3, 49) = 0.771, $p<0.516$. In other words, both groups changed in a similar pattern.

Table 17 in appendix B indicates that the assumption of Sphericity for S has been violated, which was simple responses (S) with $\chi^2 = 19.265$, $p< .002$.To obtain the accurate F values for the effect of S, tests with corrected degrees of freedom were conducted. Greenhouse-Geisser was greater than .75 so Huynh-Feldt needed to be used.

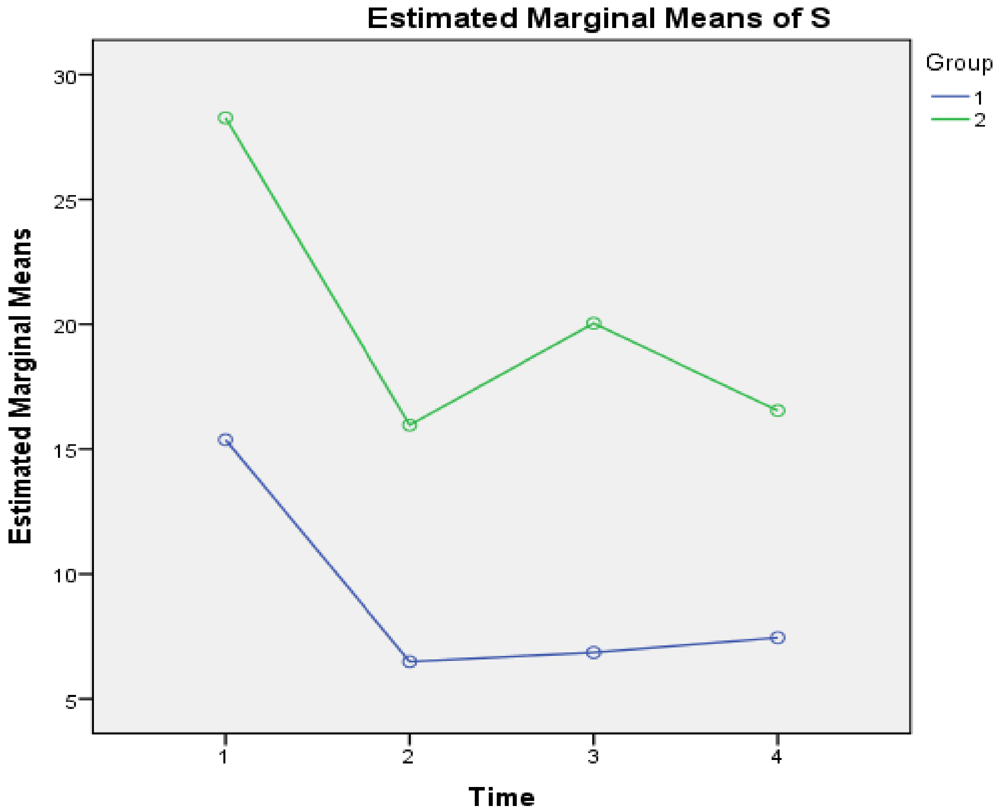
Tests with corrected degrees of freedom for S in table 18 in appendix B for the violation of Sphericity shows two hypothesis test results: (1) there was a significant difference between the two groups on S at each time point, and (2) there was no significant difference in the interaction between time and group on S at each time point.

Table 19 in appendix B shows that there was a significant difference between the two groups in terms of S, which was F (1, 51) = 26.40, $p<.001$.

Summary of lower-thinking responses

The trends and effect sizes for lower-order thinking student responses are displayed in graph 4. 10 and effect sizes.

Graph 4. 10. Trends for simple responses (S) for the treatment and control groups



Graph 4. 10 shows that there was a significant difference in simple responses (S) between the treatment and control groups. However, there was a different pattern in S between the two groups at each time point. The treatment group (group 1) had the highest frequency of S at time point 1, decreased at time point 2, and increased slightly at time points 3 and 4. The control group (group 2) had the highest frequency of S at time point 1, decreased at time point 2, increased at time point 3 and then decreased again at time point 4.

Table 4. 24. Effect size between the treatment and control group in lower-order thinking: simple responses (S)

	Time point 1	Time point 2	Time point 3	Time point 4
Between groups at each time point	-0.78 (medium)	-0.95 (large)	-1.07 (large)	-0.90 (large)

The effect sizes for student lower-order thinking (Simple responses (S)) between the two groups at each time point were calculated. The effect sizes for student lower-order thinking between the two groups at each time point started out at -0.78, decreased to -1.07 by one third until time point 3 and increased at time point 4 to -0.90. Overall, the differences between the two groups in lower-order thinking responses at different time points showed a medium to large effect size.

Summary of student responses

The differences between the two groups at each time point in higher-order thinking and lower-order thinking responses showed a large effect size. In other words, the treatment group students responded with higher-order thinking more frequently than the control group students. In contrast, the treatment group students responded with lower-order thinking less than the control group students.

The relation among teacher talk time, structure of questions (question types), and student responses

Based on the results of repeated measures ANOVA and effect sizes, relations between the three analytical categories for the study emerged. The relation between teacher talk time and the two categories (open-ended or close-ended question types and higher-order and lower-order thinking student responses) was found in both groups. The relations between individual categories were identified to evaluate how each category was related to each other statistically.

Relation between teacher talk time and a combination of question types and student responses

There was an inverse relation between teacher talk time and a combination of open-ended question types and higher-order thinking student responses. In other words, the longer teachers talked in classes, the less frequently the teachers used open-ended questions and the less frequently their students used higher-order thinking responses.

There was a direct relation between teacher talk time and a combination of close-ended question types and lower-order thinking student responses. In other words, the longer teachers talked in classes, the more frequently the teachers used close-ended questions and the more frequently their students used lower-order thinking responses.

Relation between teacher talk time and question types

There was an inverse relation between teacher talk time and open-ended question types. In other words, the longer teachers talked in classes, the less frequently the teachers used open-ended questions.

There was a direct relation between teacher talk time and close-ended question types. In other words, the longer teachers talked in classes, the more frequently the teachers used close-ended questions.

Relation between teacher talk time and student responses

There was an inverse relation between teacher talk time and higher-order thinking student responses. In other words, the longer teachers talked in classes, the less frequently their students used higher-order thinking responses.

There was a direct relation between teacher talk time and lower-order thinking student responses. In other words, the longer teachers talked in classes, the more frequently their students used lower-order thinking responses.

Relation between question types and student responses

There was a direct relation between open-ended question types and higher-order thinking student responses. In other words, the more frequently teachers used open-ended questions, the more frequently their students used higher-order thinking responses.

There was an inverse relation between open-ended question types and lower-order thinking student responses. In other words, the more frequently teachers used open-ended questions, the less frequent their students used lower-order thinking responses.

There was an inverse relation between close-ended question types and higher-order thinking student responses. In other words, the more frequently teachers used close-ended questions, the less frequent their students used higher-order thinking responses.

There was a direct relation between close-ended question types and lower-order thinking student responses. In other words, the more frequently teachers used close-ended questions, the more frequent their students used lower-order thinking responses.

Summary of the results

There were consistent differences in the three categories between the argument-based inquiry classes (treatment group) and traditional classes (control group) and within each group over the period of the project and at different time points. Table 4. 25 summarizes the patterns of the three categories between the treatment and control groups over the period of the project and at each time point.

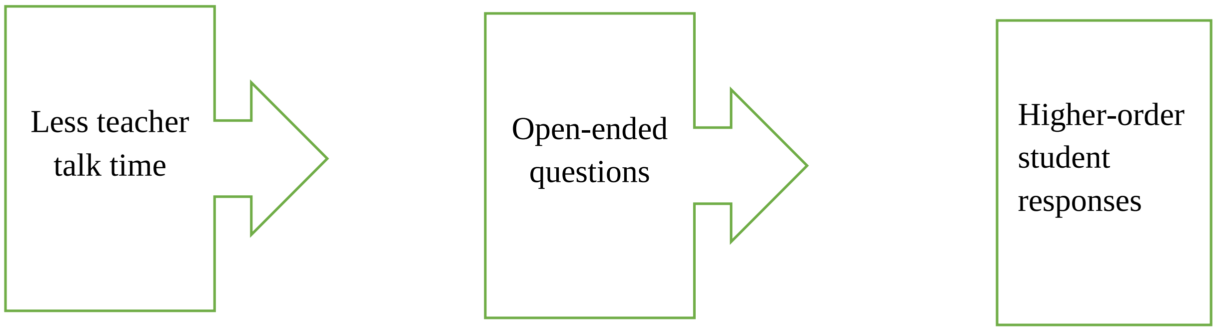
Table 4. 25. Summary of three categories between the treatment and control groups

Time Category		First semester	Second semester	Third semester	Fourth semester
Teacher talk time	Treatment	Less talk	Less talk	Less talk	Less talk
	Control	More talk	More talk	More talk	More talk
Structure of questions	Treatment	Open-ended	Open-ended	Open-ended	Open-ended
	Control	Close-ended	Close-ended	Close-ended	Close-ended
Student responses	Treatment	Higher order	Higher-order	Higher-order	Higher-order
	Control	Lower-order	Lower-order	Lower-order	Lower-order

The treatment group teachers (argument-based inquiry classes) talked less than the control group teachers, used open-ended questions more frequently than the control group teachers, and the treatment group students responded with higher-order thinking more frequently than the control group students over the period of the project and at four different time points. The control group teachers and students (traditional group) showed opposite patterns in the three categories compared to the treatment group. The control group teachers talked more than the treatment group teachers, used close-ended questions more frequently than the treatment group teachers, and the control group students responded with lower-order thinking more frequently than the control group teachers over the period of the project and at four different time points.

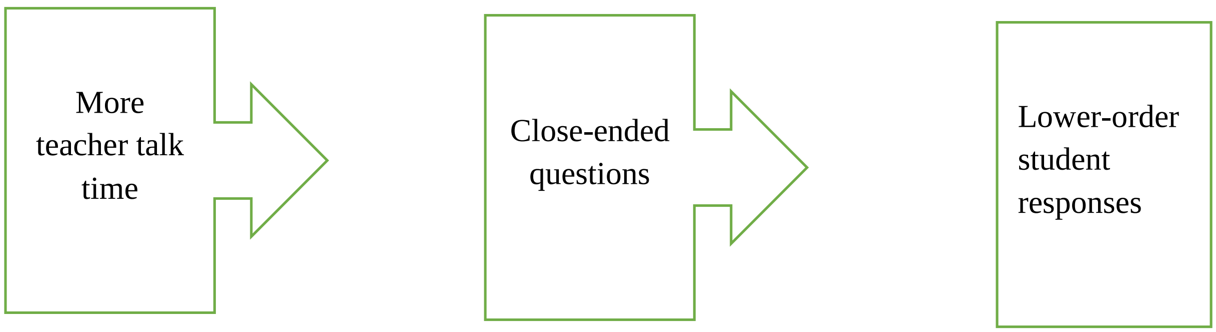
Based on these results, the relations among the three categories between the two groups emerged over the period of the project and at each time point. The relation among the three categories in the treatment group was that the length of teacher talk time was inversely related to the frequency of open-ended questions and the frequency of higher-order thinking student responses. Figure 4. 4 conceptualizes the relations of the three categories in the treatment group.

Figure 4. 4. Relations of the three categories in the treatment group



The relation of the three categories in the control group was that the length of teacher talk time was directly related to the frequency of close-ended questions and lower-order thinking student responses. Figure 4. 5 conceptualizes the relations of the three categories in the control group.

Figure 4. 5. Relations of the three categories in the control group



Students in the treatment group responded with higher-order thinking, which corresponds to the revised Bloom's taxonomy higher order thinking components (applying, analyzing, evaluating, and creating ideas) and argumentation components (claim, evidence, challenge, defend, support, and rebut). Open-ended questions encouraged students to apply, analyze, evaluate and create ideas in order to claim, challenge, defend, support and rebut ideas with evidence in their responses. Teachers also encouraged their students to talk more by asking open-ended questions which required students to use reasoning and to express their ideas without being pressured to answer correctly. In contrast, teachers in the traditional classes talked more and used more close-ended questions. That led to students responding with lower-order thinking. Close-ended questions led students to reiterate and memorize ideas rather than critique ideas. Teachers did not encourage their students to talk because longer teacher talk time and close-ended questions limited students' opportunities to express their ideas by requiring them to answer correctly. More teacher talk is related to the frequency of close-ended questions, because close-ended questions require students to respond with fixed answers for their questions without using reasoning. Students responded with lower-order thinking, which corresponds to the revised Bloom's taxonomy lower-order thinking components (remembering and understanding ideas).

While there appears to be a consistent pattern within each group, effect size calculations were used to determine the relative consistency of the pattern across the period of the study. The effect sizes for the three categories between the two groups across time (four different time points) were calculated to investigate the degree of the effect of the intervention (Cohen's d). The degrees of the effect of the engagement of the SWH approach for the treatment group teachers and students were greater than medium on average among the three categories at four different time points (refer to table 4. 26). The three categories were (1) teacher talk time, (2) structure of questions (question types), and (3) student responses. The structure of questions had two sub-categories: (1) open-ended questions and (2) close-ended questions. Under the open-ended question sub-category, there were three sub-categories: (1) asking for explanation (AE), (2) asking for self-evaluation of reasoning (AF), and (3) asking for self-evaluation of others' reasoning (AFO). Under the close-ended question sub-category, there were two sub-categories: (1) asking for factual information (AI), and (2) asking for confirmation (AC). Student responses had two sub-categories: (1) higher-order thinking and (2) lower-order thinking. Under the higher-order thinking sub-category, there were three sub-categories: (1) explanation responses (E), (2) self-evaluation of reasoning responses (SE), and (3) self-evaluation of others' reasoning responses (SEO). Under the lower-order thinking sub-category, there was one sub-category: simple responses (S).

Table 4. 26. Summary of effect sizes for the three categories

Between groups across time	Time point 1	Time point 2	Time point 3	Time point 4
Teacher Talk time	-2.81 (large)	-2.46 (large)	-1.81 (large)	-2.23 (large)
Open-ended	1.04 (large)	0.76 (medium)	0.33 (small)	0.57 (medium)
Asking for explanation (AE)	0.80 (large)	0.25 (small)	-0.02 (small)	0.46 (small)
Asking for self-evaluation of reasoning (AF)	1.16 (large)	0.62 (medium)	0.39 (small)	0.12 (small)
Asking for self-evaluation of others' reasoning (AFO)	0.84 (large)	0.86 (large)	0.60 (medium)	0.66 (medium)
Close-ended	-0.75 (medium)	-0.93 (large)	-1.07 (large)	-0.74 (medium)
Asking for information (AI)	-0.92 (large)	-2.25 (large)	-1.05 (large)	-0.71 (medium)
Asking for confirmation (AC)	-1.64 (large)	0.11 (small)	-0.32 (small)	-0.35 (small)
Higher-order	1.14 (large)	0.86 (large)	0.59 (medium)	0.76 (medium)
Explanation (E)	0.95 (large)	0.51 (medium)	0.18 (small)	0.66 (medium)
Self-evaluation of reasoning (SE)	1.09 (large)	0.56 (medium)	0.54 (medium)	0.25 (small)
Self-evaluation of others' reasoning (SEO)	0.85 (large)	0.79 (medium)	0.63 (medium)	0.60 (medium)
Lower-order: S	-0.78 (medium)	-0.95 (large)	-1.07 (large)	-0.90 (large)

Effect size was calculated from the equation for the values (i.e., means and standard deviations) of the treatment group minus the values of the control group. The effect sizes were either positive or negative depending on the category.

The effect sizes for the three categories (teacher talk time, the structure of questions, and student responses) indicate average medium degrees of effect consistently between the treatment and control group at four different time points. The patterns for teacher talk time between the two groups were opposite. The treatment group teachers consistently talked less at four different time points and the length of talk time did not fluctuate. The control group teachers talked more frequently at four different time points and the length of talk time was consistently longer than the treatment group teachers.

The patterns for the structure of questions (question types) between the two groups were different overall and at each time point. The treatment group teachers used open-ended questions the most at time point 1. A possible explanation for this pattern could be the Hawthorne effect. The treatment teachers were aware that they were participating in the project so they tried to use more open-ended questions initially. However, the frequency of open-ended questions decreased at time point 2, which may suggest that the treatment group teachers started to relapse back to their pre-treatment practice. In other words, the effect started to decrease as time passed. The frequency of open-ended questions was lower at time points 3 and 4 than at time point 1 in the treatment group but the frequency of open-ended questions was always higher in the treatment group than in the control group. The effect sizes for open-ended questions supported this finding. The effect sizes were 1.04 at time point 1, 0.76 at time point 2, 0.33 at time point 3 and 0.57 at time point 4. The effect sizes between the two groups at each time point were medium on average. This indicated that the engagement of the SWH approach impacted the treatment group teachers' practice consistently as time passed.

The difference in the frequency of asking for explanation (AE) and asking for self-evaluation of reasoning (AF) decreased as time passed in terms of the effect sizes being from large to small. However, asking for self-evaluation of others' reasoning (AFO) showed a clear difference between the treatment and control groups at each time point. This sub-type of open-ended question requires students to critique others' claims with evidence.

The patterns for close-ended questions between the two groups were the opposite of the patterns for open-ended questions. The frequency of close-ended questions for the treatment group teachers was the highest at time point 1. At time point 2, the frequency of close-ended questions in the treatment group decreased. The frequency of close-ended questions was lower at time points 3 and 4 than at time point 1. This indicated that the engagement of the SWH approach impacted the treatment group teachers on using less close-ended questions at each time point positively. The frequency of close-ended questions in the control group was highest at time point 1 and time point 4. In other words, the control group teachers appeared to show a consistent pattern throughout the duration of the study.

The frequency of higher-order thinking student responses in the treatment group was higher than the frequency of higher-order thinking student responses in the control group. The frequency of higher-order thinking student responses in the treatment group was highest at time point 1 and was directly proportional to the frequency of open-ended questions. The difference in higher-order thinking between the two groups was consistent at each time point. This result indicated that the engagement of the SWH approach consistently impacted the treatment group on providing opportunities to use higher-order thinking.

The frequency of lower-order thinking student responses in the control group was higher than in the treatment group. The frequency of lower-order thinking student responses in the treatment group was highest at time point 1. This outcome showed a relation between the frequency of close-ended questions and lower-order thinking responses. The more close-ended questions teachers asked to their students, the more lower-order thinking responses their students generated. This means teacher question types affect student response types. If teachers use open-ended questions frequently, students respond with higher-order thinking frequently. This indicates that the engagement of the SWH approach impacted the treatment group on student learning by asking questions that elicited reasoning from students.

CHAPTER 5

DISCUSSION AND IMPLICATIONS

This study examined the patterns and relations across three categories (teacher talk time, structure of questions, and student responses) between argument-based inquiry classes and traditional classes and within each group over the period of the project and at four different time points. This chapter discussed the results present in chapter 4 and provides answers to the research questions. Furthermore, findings emerged through the interpretation of the results from chapter 4. This chapter concludes with limitations and implications for the study.

Discussion of the results

There were consistent statistical differences (repeated measures ANOVA and effect sizes) between the three categories (teacher talk time, structure of questions, and student responses) between the treatment (SWH classes) and control (traditional classes) groups over the period of the project and at four different time points. An overall difference between the two groups for the three categories was indicated via statistical analysis. The differences in the three categories between the two groups emerged at four different time points. The difference in the three categories within each group emerged at different time points. In other words, there was a relation between the three categories within each group at different time points.

The effect sizes for the three categories between the two groups were greater than medium on average across time (four different time points). This indicated that the engagement of the SWH approach impacted the treatment group on the length of teacher talk time, the structure of questions and student responses to a medium degree on average.

Answers to Research Questions

The first research question was "Is there a difference in the amount of teacher talk time between teachers in argument-based inquiry classes and teachers in traditional classes?"

The answers were: (1) there was a clear difference in the amount of teacher talk time between teachers in argument-based inquiry and teachers in traditional classes over the period of the project and across time with greater than a medium effect and (2) there was a clear difference in teacher talk time within each group at each time point with greater than a medium effect.

The second research question was "Are there differences in question types and student responses between argument-based inquiry classes and traditional classes?"

The answers were (1) there was a clear difference in question types between argument-based inquiry classes and traditional classes over the period of the project and at each time point with greater than a medium effect, (2) There was a clear difference in question types within each group at each time point with greater than a medium effect, (3) There was a clear difference in student responses between argument-based inquiry classes and traditional classes over the period of the project and at each time point with greater than a medium effect, and (4) There was a clear difference in student responses within each group at each time point with greater than a medium effect.

The third question was "What are the differences in teacher talk time, teacher questioning and student responses between argument-based inquiry classes and traditional classes?"

The answers were (1) the patterns in teacher talk time between argument-based inquiry classes and traditional classes were opposite over the period of the project and at each time point. Teacher talk time in argument-based inquiry classes was less than teacher talk time in traditional classes over the period of the project and at each time point, (2) the patterns in teacher questioning between argument-based inquiry classes and traditional classes were opposite over the period of the project and at each time point. The frequency of open-ended questions in argument-based inquiry classes was higher than traditional classes over the period of the project and at each time point. The frequency of close-ended questions in argument-based inquiry classes was lower than traditional classes over the period of the project and at each time point, (3) there were three sub-type questions under open-ended questions: (a) Asking for explanation (AE), (b) Asking for self-evaluation of reasoning (AF), and (c) Asking for self-evaluation of others' reasoning (AFO). The patterns for the three sub open-ended question types between argument-based inquiry classes and traditional classes showed opposite characteristics over the period of the project and at each time point: (a) The treatment group (argument-based inquiry) teachers used Asking for explanation (AF) more frequently than the control group teachers over the period of the project and at four different time points, (b) The treatment group teachers used Asking for self-evaluation of reasoning (AF) more frequently than the control group teachers over the period of the project and at four different time points, and (c) The treatment group teachers used asking for self-evaluation of others' reasoning (AFO) more frequently than the control group teachers over the period of the project and at four different time points, (4) there were two sub close-ended question types: (1) Asking for factual information (AI), and (2) Asking for confirmation (AC). The patterns for the two sub close-ended question types between argument-based inquiry classes and traditional classes showed contrasting characteristics over the period of the project and at each time point: (1) the control group (traditional) teachers used asking for factual information (AI) more frequently than the treatment group teachers over the period of the project and at four different time points, and (2) The control group teachers used Asking for confirmation (AC) more frequently than the treatment group teachers over the period of the project and at four different time points, (5) the patterns in student responses between argument-based inquiry classes and traditional classes were opposite over the period of the project and at each time point. There were two types of student responses: (a) higher-order thinking student responses, and (b) lower-order thinking student responses. The frequency of higher-order thinking student responses in argument-based inquiry classes was higher than traditional classes over the period of the project and at each time point. The frequency of lower-order thinking student responses in argument-based inquiry classes was lower than traditional classes over the period of the project and at each time point, (6) there were three sub higher-order thinking student response types: (a) explanation responses (E), (b) self-evaluation of reasoning responses (SE), and self-evaluation of others' reasoning responses (SEO). The patterns for the three sub higher-order thinking student responses between argument-based inquiry classes and traditional classes were opposite: (a) the frequency of explanation responses (E) in argument-based inquiry classes was higher than traditional classes over the period of the project and at each time point, (b) the frequency of self-evaluation of reasoning responses (SE) in argument-based inquiry classes was higher than traditional classes over the period of the project and at each time point, and (c) the frequency of self-evaluation of others' reasoning responses (SEO) in argument-based inquiry classes was higher than traditional classes over the period of the project and at each time point, and (7) there was one sub lower-order thinking student response type: simple responses (S). The patterns for simple responses (S) between argument-based inquiry classes and traditional classes were opposite over the period of the project and at each time point. The frequency of simple responses (S) in argument-based inquiry classes was lower than traditional classes over the period of the project and at each time point.

The last question was "What is the relation between teacher talk time, teacher questioning and student responses within each group?"

The answers were (1) the relations among teacher talk time, teacher question types (open-ended questions) and student responses (higher-order thinking) in argument-based inquiry classes were a mix of inverse and directional over the period of the project and at each time point: (a) the length of teacher talk time was inversely related to a combination of the frequency of open-ended questions and the frequency of higher-order thinking student responses over the period of the project and at each time point, (b) the length of teacher talk time was inversely related to the frequency of open-ended questions over the period of the project and at each time point, (c) the length of teacher talk time was inversely related to the frequency of higher-order thinking student responses over the period of the project and at each time point, and (d) the frequency of open-ended questions was directly related to the frequency of higher-order thinking student responses over the period of the project and at each time point, and (2) the relations among teacher talk time, teacher question types (close-ended questions) and student responses (lower-order thinking) in traditional classes were directional over the period of the project and at each time point: (a) the length of teacher talk time was directly related to a combination of the frequency of close-ended questions and the frequency of lower-order thinking student responses over the period of the project and at each time point, (b) the length of teacher talk time was directly related to the frequency of close-ended questions over the period of the project and at each time point, (c) the length of teacher talk time was directly related to the frequency of lower-order thinking student responses over the period of the project and at each time point, and (d) the frequency of close-ended questions was directly related to the frequency of lower-order thinking student responses over the period of the project and at each time point.

Findings That Arose From The Results

There were four findings that emerged based on the results presented in chapter 4. The first finding was that teachers do not need to dominate discussion in an argument-based inquiry class to encourage their students to engage in the activity in terms of argumentation for learning new concepts. This finding replicated the results of previous research (Akkus, Gunel, and Hand, 2007; Fahy, 2004; Kim and Hand, 2015; Rowe, 1974**)** that showed that the length of teacher talk time was inversely correlated to the length of student talk time. In an argument-based inquiry class, teachers enable students to argue about their claims with evidence. In other words, student talk time is encouraged instead of teacher talk time. In a traditional class, teachers take the lead in class by asking many close-ended questions that require students to respond with specific answers. In this approach, students may be deprived from expressing their opinions about the topics that they are learning. In this type of class, students may not be provided opportunities to use reasoning and instead feel pressured to respond with specific answers.

The second finding was that teachers should be encouraged to use open-ended questions rather than close-ended questions to promote student reasoning in argumentation. This finding replicated the results of previous research regarding the effect of open-ended questions on student cognitive levels in terms of student responses requiring higher-order thinking (Berland & Reiser, 2009; Chin, 2007; McNeill & Pimentel, 2009; NRC, 2008; Suppe, 1998). The results from this study showed that in the argument-based inquiry classes, teachers used open-ended questions more frequently than traditional classes. The frequency of open-ended questions is directly related to the frequency of student higher-order thinking responses. Open-ended questions promote student use of argumentation components (claim, evidence, challenge, support, defend, and rebut) because open-ended questions are designed to elicit explanations, self-evaluation of reasoning and self-evaluation of others' reasoning about ideas from student responses, which all require higher-order thinking. In other words, open-ended questions are vehicles for helping students move up to the next cognitive level within the Zone of Proximal Development [ZPD], so the students are able to learn new concepts.

The third finding was that students should be encouraged to respond with higher-order thinking reflective of Bloom's revised taxonomy and argumentation components. This finding replicated the results of previous research regarding student responses (Anderson et al, 2001; Barnes, 1976; Britton, 1982; Bruner, 1986; Corson, 1988; Kim and Hand, 2015; Lemke, 1990; Maloch, 2002; NRC, 2012; Pea, 1993; Prawat, 1993; Rivard & Straw, 2000; Schoenfeld, 1989). The more open-ended questions that were asked, the more students responded with higher-order thinking. The more close-ended questions that were asked, the more lower-order thinking responses were generated from students. These patterns indicated that teachers should be encouraged to use open-ended questions in order for students to generate responses with higher-order thinking (reasoning). Student responses with higher-order thinking showed students moved up to the next cognitive level within the ZPD.

The fourth finding was that teachers should be encouraged to adopt an appropriate combination of shortening teacher talk time and increasing the frequency of open-ended questions in order to elicit student higher-order thinking responses. The relations among the three categories (teacher talk time, structure of questions, and student responses) within each group indicated that teachers are encouraged to use the two categories (teacher talk time and structure of questions) in an appropriate manner. This enables students to use reasoning more frequently to generate higher-order thinking responses. To be effective for student learning, teachers are encouraged to scaffold their questions associated with talk time within the ZPD.

The fourth finding extended previous research presented in chapter 2 regarding the effect of open-ended questions on student learning in terms of types of student response via a further evaluation of teacher question types in relation to student responses by providing students opportunities to critique ideas (e.g., reduced teacher talk time contributes to increased student talk time). The extension begins with the assertion that there are not enough empirical studies on the ways that open-ended questioning plays a role in students' cognition and learning (Lee & Kinzie, 2012). Furthermore, even though many researchers concluded that teachers' open-ended questions that focus on promoting higher-cognitive levels correlate to students' responses requiring higher-order thinking, some researchers questioned this finding by asserting that there is a low correlation between the cognitive level of teachers' questions and the cognitive level of students' answers (e.g., Dillon 1982b; Mills et al. 1980). This disagreement between researchers on the impact of open-ended questions on leading to higher-order thinking student responses requires more study. The research from this study suggest that there is a relationship between student talk time and the structure of questions, with clear evidence of a difference between the treatment (SWH classes) and control groups (traditional classes. This extension was supported by the results of studies conducted by McNeill and Pimentel (2010) and Martin and Hand (2009). They explained the effect of open-ended questions on student learning in argument-based inquiry classes by asserting that there was a positive relation between a higher frequency of open-ended questions from teachers, a higher percentage of student talk, a higher frequency of using reasoning in terms of providing evidence to support their claims, and a higher interaction between students in terms of argumentation. In other words, when teachers did not dominate talk, students were provided opportunities to critique ideas that were discussed during the classes, utilizing argumentation components.

Kim and Hand (2015) also provided a direction for the study by suggesting that teacher question types that are able to facilitate student reasoning processes in an argument-based inquiry class need to be studied based on their analysis of teacher and student discourse patterns between argument-based inquiry classes and traditional classes. Their analysis results showed that the argument-based inquiry class teachers and students utilized critique components (challenge, defense, support and reject ideas), which are crucial for argumentation, more frequently than the traditional class teachers and students. These results indicated that teacher question types were directly related to student responses while learning new concepts. Furthermore, based on the results from the initial analyses, any relations among teacher talk time, structure of question, and student response warranted further investigation. This study' questions were expanded to include an evaluation of the relations between the three categories (teacher talk time, the structure of questions and student responses) based on this suggestion.

Table 5. 1 shows the relations between the current study and previous research in terms of conceptualizing the associations between the findings from the study and previous research.

Table 5. 1. Summary of relations between the study and previous research

Directions	Findings	Previous research	Relations
Studied previously	Teacher talk time	Teacher talk time	Replicate
Studied previously	Structure of questions	Question types	Replicate
Studied previously	Student responses	Cognitive levels	Replicate
Needs more study	Combination of the three categories	Correlation between teacher question types and student responses	Extension

Previous research has investigated teacher talk time, the structure of questions, and student responses. This study replicated the results of these three categories as the stepping stone for extending this study. There is limited research about the combination of the two categories (teacher talk and the structure of questions) in order to promote student higher-order thinking responses. Based on the replication of the results of the three categories, this study's questions were extended to calculate the correlations between teacher talk time and the structure of questions to find an appropriate manner to elicit student reasoning processes.

The findings of the study add to the role of teachers in argument-based inquiry classes explored in previous research. Chin and Osborne (2008), Mortimer and Scott (2003), and King (1996) elucidated the role of teachers in inquiry-based classes by emphasizing the fact that the core of teaching science is to associate students with the dialogical practices of science (e.g., teachers used open-ended questions frequently to promote student higher-order thinking responses by guaranteeing student talk time). The teacher must help students to link scientific concepts to a social context and encourage students to develop the ability to apply the concepts to diverse situations on their own. In doing so, the students are eventually able to internalize and make sense of the new concepts. Teachers should assist students in utilizing higher cognitive components (i.e., Bloom's revised taxonomy) by creating learning supportive environments where open-ended questions are asked frequently, teacher talk time is not dominant and students are provided opportunities to express their ideas without feeling pressured to respond with fixed answers.

Limitations of the study

There were four limitations to the current study. First, the study used data that was already collected from the SWH project. For example, teachers' talk time was measured in seconds from the project, structure of questions (question types), and student responses were counted numerically. There were no interactions between the researcher and the teachers in both groups and thus the study relied on previously collected data so the researcher was not able to measure additional factors that may have affected the teachers and students' performance such as teachers' content knowledge, teaching experience, and student achievement.

Second, there were some factors the study was not able to control. The study was not able to control that the two groups had the same baseline when the researcher was choosing the participants of the study. For example, the study couldn't dictate that teachers taught the same subject/topic during the same semester and measure that the teachers and students in each group had the same or a similar level of content knowledge. The study used the convenient sampling method since the samples were already collected from the project. The study was not able to identify and quantify exactly how much the SWH approach (argument-based inquiry) effected student learning between the treatment and control groups in terms of student response types in relation to teacher question types and teacher talk time. The study was able to discover the effect of the SWH approach between the two groups on student learning regardless of what subject the teachers taught and the level of content knowledge the teachers and students possessed, and relations between the three categories (teacher talk time, teacher question types and student responses).

Third, there was no direct experimental manipulation in the study. For this reason, this study was not able to identify causal relations between the three categories. The study was able to tell us that there were relations among the three categories. The causal relations among the three categories need to be identified so when teachers design their lesson plans, those factors may be considered to promote student reasoning in terms of encouraging student higher-order thinking responses. In other words, teachers understand what types of questions and what length of talk time they have to use to elicit a certain type of student response.

Finally, due to the limitations on the usage of the data, question sequences were not available to be analyzed. Previous research (e.g., Chin, 2007) has suggested that a series of questions is important for student learning since it can guide students to engage in argumentation appropriately. However, the current study was not able to address this issue.

Implications

The engagement of the SWH approach impacted the treatment group teachers and students in terms of reduced teacher talk time, a higher frequency of open-ended questions, and a higher frequency of student higher-order thinking responses compared to the control group teachers and students. However, the frequency of open-ended questions for the treatment group teachers was at the highest at time point 1. The frequency of open-ended questions was lower across the rest of time (time points 2, 3 and 4). This result suggested that more study is needed in understanding how to maintain teachers' use of open-ended questions at each time point. The frequency of student higher-order thinking responses was directly related to the frequency of open-ended questions. This means that teachers should be encouraged to pose questions that promote student reasoning in terms of argumentation components and Bloom's revised taxonomy components. In addition, teachers are also encouraged to generate a series of questions that can elicit student reasoning.

The length of teacher talk time in the treatment group teachers suggested that when teachers talked less in class students engaged in argumentation more in terms of responding with higher-order thinking instead of lower-order thinking, compared to when teacher talk dominated the class time. In contrast, when teacher talk dominated the class time, students responded more frequently with lower-order thinking than higher-order thinking. Additional research needs to address relations between teacher talk time and student talk time in promoting argumentation.

The appropriate length of teacher talk time may help create supportive learning environments for student learning when it provides students with more opportunities to express and exchange ideas through critiquing.

There were four future directions for the study. First, the relations among the three categories within each group were identified. This study was not experimental. Causal relations among the three categories need to be studied in detail for future research to provide explanations of how open-ended questions affect student learning directly in terms of promoting reasoning (higher-order thinking).

Second, the methods to sustain teachers' use of open-ended questions across an academic year need to be investigated further to make the SWH approach effective for student learning. The frequency of open-ended questions increased during the period of the project because the teachers were aware that they were in the project and were being observed and participated in PD sessions. It is important for teachers to recognize the importance of open-ended questions and use them appropriately. In order for teachers to use open-ended questions effectively, methods courses should emphasize the importance and usefulness of open-ended questions for student learning and encourage teachers to use open-ended questions consistently.

Third, the relation between teacher talk time and student learning in terms of the degree of engagement in argumentation activities needs to be studied in order to persuade teachers to shift their roles from traditional to argument-based inquiry practice. When teachers are asked to use inquiry, they may feel that their role is decreased in terms of reduced talk time (NRC, 1996). Future studies should focus on ways of benefiting student learning by relinquishing teacher talk time to student talk time in class. The role of teachers in class is to help their students learn concepts by encouraging them to use argumentation components. It is a teacher's job to create learning supportive environments where students are encouraged to talk about ideas critically by providing adequate student talk time.

Fourth, question sequences need to be investigated in detail. What type of question sequence affects student learning the most needs to be evaluated. The structure of questioning that elicits student reasoning should be identified. Teachers are recommended to design a series of questions that considers their students' needs. They need to be proficient at using question sequences that are appropriate to the learning situation and student level of content knowledge. In this way, student learning will be increased in argument-based inquiry classes.

APPENDIX A. CODE BOOK FOR TEACHER QUESTIONS AND STUDENT RESPONSES

Structure of questions

Table A. 1. Codes for structure of questions

Type	Code	Example
Open-ended question (**O**)	Asking for explanation based on experience, evidence or data (**AE**)	**O/AE** - How does air gets into you compared to your model?
	Asking for self-evaluation of reasoning (**AF**)	**O/AF** - How did you make it different/better?
	Asking for evaluation of other's reasoning (**AFO**)	**O/AFO** - What do you think (about his idea)?
Close-ended question (**C**)	Asking for factual information (**AI**)	**C/AI** - When we do research, we need an expert's help. Does that help?
	Asking for confirmation (**AC**)	**C/AC** - Does your model work the way the human body works?

Student response

Table A. 2. Codes or code for student response

Type	Code	Example
Lower-order thinking (**L**)	Simple response (**S**)	**L/S** - Have you seen models? Yes.
Higher-order thinking (**H**)	Explanation response (**E**)	**H/E** - How did the air get into the lungs? When we breathe in, air comes into us.
	Self-evaluation of one's own idea response (**SE**)	**H/SE** - Does your model work the way the human body works? No, it didn't work as I expected, I missed some parts in my model.
	Self-evaluation of others' ideas response (**SEO**)	**H/SEO** - Why did you do that?

APPENDIX B. REPEATED MEASURES ANOVA RESULTS FOR THREE CATEGORIES

Repeated measures ANOVA for Teacher talk time

Table B. 1. Maulchly's test of Sphericit for teacher talk time

Measure: length

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Talktime	.785	12.026	5	.034	.884	.955	.333

Table B. 2. Tests with corrected degrees of freedom for a violation of Sphericity for teacher talk time

Measure: Length

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Talktime	Sphericity Assumed	1846.173	3	615.391	6.605	.000
	Greenhouse-Geisser	1846.173	2.653	696.002	6.605	.001
	Huynh-Feldt	1846.173	2.866	644.067	6.605	.000
	Lower-bound	1846.173	1.000	1846.173	6.605	.013
Talktime * Group	Sphericity Assumed	723.599	3	241.200	2.589	.055
	Greenhouse-Geisser	723.599	2.653	272.794	2.589	.063
	Huynh-Feldt	723.599	2.866	252.439	2.589	.058
	Lower-bound	723.599	1.000	723.599	2.589	.114
Error(Talktime)	Sphericity Assumed	14255.740	153	93.175		

	Greenhouse-Geisser	14255.740	135.280	105.380
	Huynh-Feldt	14255.740	146.188	97.516
	Lower-bound	14255.740	51.000	279.524

Table B. 3. Tests of Between-Subjects Effects for teacher talk time

Measure: Length

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	33582.082	1	33582.082	445.592	.000
Group	9414.662	1	9414.662	124.921	.000
Error	3843.623	51	75.365		

Repeated measures ANOVA for open-ended questions

Table B. 4. Multivariate Analysis of Variance (MANOVA) for open-ended questions

Effect			Value	F	Hypothesis df	Error df	Sig.
Between Subjects	Intercept	Pillai's Trace	.729	44.019^b	3.000	49.000	.000
		Wilks' Lambda	.271	44.019^b	3.000	49.000	.000
		Hotelling's Trace	2.695	44.019^b	3.000	49.000	.000
		Roy's Largest Root	2.695	44.019^b	3.000	49.000	.000
	Group	Pillai's Trace	.360	9.200^b	3.000	49.000	.000
		Wilks' Lambda	.640	9.200^b	3.000	49.000	.000
		Hotelling's Trace	.563	9.200^b	3.000	49.000	.000
		Roy's Largest Root	.563	9.200^b	3.000	49.000	.000

Within Subjects	Time	Pillai's Trace	.369	2.800[b]	9.000	43.000	.011
		Wilks' Lambda	.631	2.800[b]	9.000	43.000	.011
		Hotelling's Trace	.586	2.800[b]	9.000	43.000	.011
		Roy's Largest Root	.586	2.800[b]	9.000	43.000	.011
	Time * Group	Pillai's Trace	.354	2.613[b]	9.000	43.000	.017
		Wilks' Lambda	.646	2.613[b]	9.000	43.000	.017
		Hotelling's Trace	.547	2.613[b]	9.000	43.000	.017
		Roy's Largest Root	.547	2.613[b]	9.000	43.000	.017

Table B. 5. Maulchly's test of Sphericity for open-ended questions

Within Subjects Effect	Measure	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
						Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	AE	.570	27.958	5	.000	.722	.770	.333
	AF	.281	63.145	5	.000	.547	.574	.333
	AFO	.573	27.708	5	.000	.719	.767	.333

Table B. 6. Univariate Tests for open-ended sub question types

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	AE	Sphericity Assumed	323.104	3	107.701	3.233	.024
		Greenhouse-Geisser	323.104	2.166	149.174	3.233	.039
		Huynh-Feldt	323.104	2.310	139.885	3.233	.036
		Lower-bound	323.104	1.000	323.104	3.233	.078
	AF	Sphericity Assumed	175.328	3	58.443	4.834	.003
		Greenhouse-Geisser	175.328	1.642	106.802	4.834	.015
		Huynh-Feldt	175.328	1.722	101.804	4.834	.014
		Lower-bound	175.328	1.000	175.328	4.834	.032
	AFO	Sphericity Assumed	60.784	3	20.261	1.655	.179
		Greenhouse-Geisser	60.784	2.158	28.162	1.655	.194
		Huynh-Feldt	60.784	2.301	26.415	1.655	.191

		Lower-bound	60.784	1.000	60.784	1.655	.204
Time * Group	AE	Sphericity Assumed	513.311	3	171.104	5.136	.002
		Greenhouse-Geisser	513.311	2.166	236.991	5.136	.006
		Huynh-Feldt	513.311	2.310	222.234	5.136	.005
		Lower-bound	513.311	1.000	513.311	5.136	.028
	AF	Sphericity Assumed	334.875	3	111.625	9.234	.000
		Greenhouse-Geisser	334.875	1.642	203.992	9.234	.001
		Huynh-Feldt	334.875	1.722	194.445	9.234	.000
		Lower-bound	334.875	1.000	334.875	9.234	.004
	AFO	Sphericity Assumed	75.916	3	25.305	2.067	.107
		Greenhouse-Geisser	75.916	2.158	35.173	2.067	.128
		Huynh-Feldt	75.916	2.301	32.990	2.067	.124
		Lower-bound	75.916	1.000	75.916	2.067	.157
Error(Time)	AE	Sphericity Assumed	5096.972	153	33.314		
		Greenhouse-Geisser	5096.972	110.464	46.142		
		Huynh-Feldt	5096.972	117.798	43.269		
		Lower-bound	5096.972	51.000	99.941		
	AF	Sphericity Assumed	1849.615	153	12.089		
		Greenhouse-Geisser	1849.615	83.722	22.092		
		Huynh-Feldt	1849.615	87.833	21.058		
		Lower-bound	1849.615	51.000	36.267		
	AFO	Sphericity Assumed	1873.150	153	12.243		
		Greenhouse-Geisser	1873.150	110.076	17.017		
		Huynh-Feldt	1873.150	117.360	15.961		
		Lower-bound	1873.150	51.000	36.728		

Table B. 7. Tests of between-subject effects for open-ended questions

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	AE	1767.891	1	1767.891	127.485	.000
	AF	324.748	1	324.748	61.979	.000
	AFO	270.770	1	270.770	47.555	.000
Group	AE	95.679	1	95.679	6.900	.011
	AF	87.276	1	87.276	16.657	.000
	AFO	108.912	1	108.912	19.128	.000
Error	AE	707.241	51	13.867		
	AF	267.224	51	5.240		

	AFO	290.388	51	5.694

Repeated measures ANOVA for close-ended questions

Table B. 8. Multivariate Analysis of Variance (MANOVA) for close-ended question

Effect			Value	F	Hypothesis df	Error df	Sig.
Between Subjects	Intercept	Pillai's Trace	.782	89.910[b]	2.000	50.000	.000
		Wilks' Lambda	.218	89.910[b]	2.000	50.000	.000
		Hotelling's Trace	3.596	89.910[b]	2.000	50.000	.000
		Roy's Largest Root	3.596	89.910[b]	2.000	50.000	.000
	Group	Pillai's Trace	.338	12.779[b]	2.000	50.000	.000
		Wilks' Lambda	.662	12.779[b]	2.000	50.000	.000
		Hotelling's Trace	.511	12.779[b]	2.000	50.000	.000
		Roy's Largest Root	.511	12.779[b]	2.000	50.000	.000
Within Subjects	Time	Pillai's Trace	.386	4.823[b]	6.000	46.000	.001
		Wilks' Lambda	.614	4.823[b]	6.000	46.000	.001
		Hotelling's Trace	.629	4.823[b]	6.000	46.000	.001
		Roy's Largest Root	.629	4.823[b]	6.000	46.000	.001
	Time * Group	Pillai's Trace	.244	2.472[b]	6.000	46.000	.037
		Wilks' Lambda	.756	2.472[b]	6.000	46.000	.037
		Hotelling's Trace	.322	2.472[b]	6.000	46.000	.037
		Roy's Largest Root	.322	2.472[b]	6.000	46.000	.037

Table B. 9. Maulchly's test of Sphericity for close-ended questions

Within Subjects Effect	Measure	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
						Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	AI	.760	13.650	5	.018	.843	.907	.333
	AC	.807	10.634	5	.059	.870	.939	.333

Table B. 10. Univariate Tests for close-ended questions

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	AI	Sphericity Assumed	2946.643	3	982.214	8.431	.000
		Greenhouse-Geisser	2946.643	2.528	1165.823	8.431	.000
		Huynh-Feldt	2946.643	2.722	1082.393	8.431	.000
		Lower-bound	2946.643	1.000	2946.643	8.431	.005
	AC	Sphericity Assumed	146.868	3	48.956	3.623	.015
		Greenhouse-Geisser	146.868	2.611	56.259	3.623	.019
		Huynh-Feldt	146.868	2.818	52.119	3.623	.017
		Lower-bound	146.868	1.000	146.868	3.623	.063
Time * Group	AI	Sphericity Assumed	623.549	3	207.850	1.784	.153
		Greenhouse-Geisser	623.549	2.528	246.704	1.784	.162
		Huynh-Feldt	623.549	2.722	229.049	1.784	.158
		Lower-bound	623.549	1.000	623.549	1.784	.188
	AC	Sphericity Assumed	92.849	3	30.950	2.290	.081
		Greenhouse-Geisser	92.849	2.611	35.567	2.290	.090
		Huynh-Feldt	92.849	2.818	32.949	2.290	.085
		Lower-bound	92.849	1.000	92.849	2.290	.136
Error(Time)	AI	Sphericity Assumed	17824.697	153	116.501		

		Greenhouse-Geisser	17824.697	128.904	138.279
		Huynh-Feldt	17824.697	138.839	128.384
		Lower-bound	17824.697	51.000	349.504
	AC	Sphericity Assumed	2067.386	153	13.512
		Greenhouse-Geisser	2067.386	133.138	15.528
		Huynh-Feldt	2067.386	143.716	14.385
		Lower-bound	2067.386	51.000	40.537

Table B. 11. Tests of Between-Subject Effects for close-ended questions

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	AI	10110.881	1	10110.881	154.254	.000
	AC	589.497	1	589.497	60.019	.000
Group	AI	1701.069	1	1701.069	25.952	.000
	AC	4.436	1	4.436	.452	.505
Error	AI	3342.898	51	65.547		
	AC	500.911	51	9.822		

Repeated measures ANOVA for higher-order student responses

Table B. 12. Multivariate Analysis of Variance (MANOVA) for higher-order thinking

Effect			Value	F	Hypothesis df	Error df	Sig.
Between Subjects	Intercept	Pillai's Trace	.724	42.796[b]	3.000	49.000	.000
		Wilks' Lambda	.276	42.796[b]	3.000	49.000	.000
		Hotelling's Trace	2.620	42.796[b]	3.000	49.000	.000

		Roy's Largest Root	2.620	42.796[b]	3.000	49.000	.000
	Group	Pillai's Trace	.384	10.166[b]	3.000	49.000	.000
		Wilks' Lambda	.616	10.166[b]	3.000	49.000	.000
		Hotelling's Trace	.622	10.166[b]	3.000	49.000	.000
		Roy's Largest Root	.622	10.166[b]	3.000	49.000	.000
Within Subjects	Time	Pillai's Trace	.292	1.968[b]	9.000	43.000	.067
		Wilks' Lambda	.708	1.968[b]	9.000	43.000	.067
		Hotelling's Trace	.412	1.968[b]	9.000	43.000	.067
		Roy's Largest Root	.412	1.968[b]	9.000	43.000	.067
	Time * Group	Pillai's Trace	.336	2.421[b]	9.000	43.000	.025
		Wilks' Lambda	.664	2.421[b]	9.000	43.000	.025
		Hotelling's Trace	.507	2.421[b]	9.000	43.000	.025
		Roy's Largest Root	.507	2.421[b]	9.000	43.000	.025

Table B.13. Maulchly's test of Sphericity for higher-order thinking responses

Within Subjects Effect	Measure	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
						Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	E	.655	21.059	5	.001	.765	.819	.333
	SE	.334	54.499	5	.000	.582	.613	.333
	SEO	.750	14.316	5	.014	.842	.907	.333

Table B. 14. Univariate Tests for higher-order thinking student responses

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	E	Sphericity Assumed	223.738	3	74.579	2.875	.038
		Greenhouse-Geisser	223.738	2.295	97.483	2.875	.053
		Huynh-Feldt	223.738	2.456	91.080	2.875	.049
		Lower-bound	223.738	1.000	223.738	2.875	.096
	SE	Sphericity Assumed	85.705	3	28.568	3.249	.024
		Greenhouse-Geisser	85.705	1.747	49.049	3.249	.050
		Huynh-Feldt	85.705	1.840	46.587	3.249	.047
		Lower-bound	85.705	1.000	85.705	3.249	.077
	SEO	Sphericity Assumed	16.940	3	5.647	.626	.599
		Greenhouse-Geisser	16.940	2.527	6.703	.626	.572
		Huynh-Feldt	16.940	2.722	6.224	.626	.584
		Lower-bound	16.940	1.000	16.940	.626	.432
Time * Group	E	Sphericity Assumed	434.757	3	144.919	5.586	.001
		Greenhouse-Geisser	434.757	2.295	189.424	5.586	.003
		Huynh-Feldt	434.757	2.456	176.982	5.586	.003
		Lower-bound	434.757	1.000	434.757	5.586	.022
	SE	Sphericity Assumed	183.592	3	61.197	6.960	.000
		Greenhouse-Geisser	183.592	1.747	105.069	6.960	.002
		Huynh-Feldt	183.592	1.840	99.796	6.960	.002
		Lower-bound	183.592	1.000	183.592	6.960	.011
	SEO	Sphericity Assumed	21.053	3	7.018	.779	.508
		Greenhouse-Geisser	21.053	2.527	8.331	.779	.488
		Huynh-Feldt	21.053	2.722	7.735	.779	.497
		Lower-bound	21.053	1.000	21.053	.779	.382

Error(Time)	E	Sphericity Assumed	3969.064	153	25.942
		Greenhouse-Geisser	3969.064	117.053	33.908
		Huynh-Feldt	3969.064	125.281	31.681
		Lower-bound	3969.064	51.000	77.825
	SE	Sphericity Assumed	1345.267	153	8.793
		Greenhouse-Geisser	1345.267	89.114	15.096
		Huynh-Feldt	1345.267	93.823	14.338
		Lower-bound	1345.267	51.000	26.378
	SEO	Sphericity Assumed	1379.211	153	9.014
		Greenhouse-Geisser	1379.211	128.881	10.701
		Huynh-Feldt	1379.211	138.813	9.936
		Lower-bound	1379.211	51.000	27.043

Table B. 15. Tests of Between-Subjects Effects for higher-order thinking student responses

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	E	1475.388	1	1475.388	121.195	.000
	SE	196.772	1	196.772	48.239	.000
	SEO	153.141	1	153.141	41.094	.000
Group	E	174.633	1	174.633	14.345	.000
	SE	65.262	1	65.262	15.999	.000
	SEO	69.151	1	69.151	18.556	.000
Error	E	620.855	51	12.174		
	SE	208.033	51	4.079		

	SEO	190.057	51	3.727

Repeated measures ANOVA for lower-order thinking student responses

Table B. 16. Multivariate Analysis of Variance (MANOVA) for lower-order thinking

Effect		Value	F	Hypothesis df	Error df	Sig.
Time	Pillai's Trace	.294	6.808[b]	3.000	49.000	.001
	Wilks' Lambda	.706	6.808[b]	3.000	49.000	.001
	Hotelling's Trace	.417	6.808[b]	3.000	49.000	.001
	Roy's Largest Root	.417	6.808[b]	3.000	49.000	.001
Time * Group	Pillai's Trace	.045	.771[b]	3.000	49.000	.516
	Wilks' Lambda	.955	.771[b]	3.000	49.000	.516
	Hotelling's Trace	.047	.771[b]	3.000	49.000	.516
	Roy's Largest Root	.047	.771[b]	3.000	49.000	.516

Table B. 17. Maulchly's test of Sphericity for lower-order thinking responses

Measure: S

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	.679	19.265	5	.002	.785	.841	.333

Table B. 18. Tests of Within-Subjects Effects for lower-order thinking student responses

Measure: S

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.
Time	Sphericity Assumed	3797.496	3	1265.832	9.759	.000

	Greenhouse-Geisser	3797.496	2.355	1612.817	9.759	.000
	Huynh-Feldt	3797.496	2.524	1504.416	9.759	.000
	Lower-bound	3797.496	1.000	3797.496	9.759	.003
Time * Group	Sphericity Assumed	188.364	3	62.788	.484	.694
	Greenhouse-Geisser	188.364	2.355	79.999	.484	.648
	Huynh-Feldt	188.364	2.524	74.622	.484	.661
	Lower-bound	188.364	1.000	188.364	.484	.490
Error(Time)	Sphericity Assumed	19844.750	153	129.704		
	Greenhouse-Geisser	19844.750	120.083	165.258		
	Huynh-Feldt	19844.750	128.736	154.151		
	Lower-bound	19844.750	51.000	389.113		

Table B. 19. Tests of Between-Subjects Effects for lower-order thinking student responses

Measure: S

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	11323.615	1	11323.615	181.066	.000
Group	1651.087	1	1651.087	26.401	.000
Error	3189.465	51	62.539		

APPENDIX C. PARTICIPANTS DEMOGRAPHICS

Table C. 1. Gender distribution in the treatment and control groups

Code (Treatment)	Gender	Code (Control)	Gender
1-0141-1	F	2-2231-2	F
1-6041-4	F	2-5131-4	F
1-6052-4	F	2-5334-4	F
1-6241-4	F	2-5831-4	M
1-0951-1	F	2-5942-4	F
1-0952-1	F	2-1451-1	F
1-5631-4	F	2-2531-2	F
1-0143-5	F	2-2541-2	F
1-5751-4	F	2-5141-4	F
1-6045-4	F	2-5351-4	M
1-6050-4	F	2-5151-4	F
1-6131-4	F	2-2441-2	F
1-0261-1	F	2-5336-4	F
1-0732-1	F	2-5353-4	F
1-2732-2	F	2-5841-4	F
1-0231-1	F	2-6351-4	F
1-0741-1	F	2-1351-1	F
1-0953-1	F	2-6331-4	F
1-3631-3	F	2-3151-3	F
1-3641-3	F	2-3531-3	F
1-5643-4	F	2-5332-4	F
1-5761-4	F	2-5343-4	F
1-6039-4	F	2-5354-4	F
1-6046-4	F	2-2431-2	F
1-6051-4	F	2-3933-3	F
1-6057-4	F	2-5333-4	M
1-6142-4	M	2-1541-1	F
1-2751-2	F	2-4141-3	F
1-6031-4	F	2-4131-3	F
1-6032-4	F	2-1531-1	F
Ratio (M/F)	1/29 (3%)	Ratio (M/F)	3/27 (11%)

M: male, F: female. The ratios between the number of males and the number of females in both groups were calculated in percentages by dividing the number of males by the number of females.

Table C. 2. Teacher talk time in the treatment group at each time point

Clip Code	First year (Fall)	First year (Spring)	Second year (Fall)	Second year (Spring)
1-0141-1	01:15/21:28	00:14/21:06	00:30/15:55	00:23/19:38
1-6041-4	08:04/25:07	02:05/15:03	00:28/09:28	00:06/05:57

1-6052-4	01:32/21:26	05:13/12:43	13:12/34:48	06:03/24:33
1-6241-4	06:23/33:31	03:25/21:56	02:28/26:21	01:35/15:07
1-0951-1	00:00/22:54	00:55/17:34	02:46/01:01:46	00:37/43:12
1-0952-1	01:11/26:19	07:06/01:04:11	00:44/31:19	Missing
1-5631-4	06:28/31:10	03:00/27:12	03:33/22:07	02:19/19:11
1-0143-5	00:30/15:45	00:44/10:19	01:01/16:13	Missing
1-5751-4	02:50/23:57	01:22/10:58	00:41/09:44	00:49/10:35
1-6045-4	12:11/50:00	01:37/14:36	01:11/10:19	01:30/11:15
1-6050-4	04:17/01:05:32	00:12/16:32	00:13/10:03	00:15/12:26
1-6131-4	10:45/33:19	02:08/13:48	02:29/16:42	01:07/15:27
1-0261-1	00:43/26:31	02:26/35:42	03:50/11:59	01:50/01:41:35
1-0732-1	03:37/12:24	00:50/29:14	01:35/30:09	03:09/14:09
1-2732-2	08:20/35:22	14:32/29:34	12:26/30:21	06:40/19:26
1-0231-1	00:00/19:19	00:12/32:04	00:00/12:14	00:09/24:16
1-0741-1	03:32/30:40	00:15/07:08	00:24/28:31	00:33/16:46
1-0953-1	08:08/01:12:47	01:47/09:55	03:07/29:48	05:21/39:49
1-3631-3	07:07/01:45:59	00:58/19:30	00:07/02:00	02:25/22:48
1-3641-3	04:35/22:08	02:16/38:17	00:34/38:03	Missing
1-5643-4	06:37/25:16	04:28/39:27	03:17/17:32	01:08/06:17
1-5761-4	01:14/10:01	1:05/13:33	01:04/11:11	02:41/20:58
1-6039-4	01:00/33:16	01:24/14:57	00:22/05:49	01:25/11:49
1-6046-4	01:24/15:49	00:25/07:43	00:26/07:14	01:01/11:59
1-6051-4	03:52/18:29	03:09/13:01	00:13/12:36	00:16/13:30
1-6057-4	03:10/29:05	01:35/06:23	01:42/15:33	00:06/07:53
1-6142-4	06:27/35:55	02:18/22:27	00:00/11:50	00:13/19:17
1-2751-2	05:10/27:04	00:55/08:06	05:28/21:50	06:43/27:23
1-6031-4	04:06/27:20	00:17/11:19	01:27/15:16	01:02/20:08
1-6032-4	00:34/16:05	00:11/18:23	01:42/13:00	00:04/06:19
Average percentage	14.17	11.53	10.61	9.52

Time measured in seconds and averaged time converted into percentage at each time point

Table C. 3. Teacher talk time in the control group at each time point

Code	Clip	First year (Fall)	First year (Spring)	Second year (Fall)	Second year (Spring)
2-2231-2		03:27/05:18	11:45/27:43	09:55/15:46	Missing
2-5131-4		05:13/15:16	03:26/11:33	02:13/07:19	06:16/16:48
2-5334-4		13:15/23:44	15:24/20:41	05:05/20:05	08:02/19:27
2-5831-4		13:46/35:20	13:04/20:42	16:16/32:20	12:16/28:34
2-5942-4		13:01/29:58	05:16/25:26	08:08/37:25	09:26/27:34
2-1451-1		16:07/42:24	05:44/18:31	02:10/28:33	05:03/23:29
2-2531-2		18:22/42:41	17:36/38:45	11:21/21:29	10:39/28:50
2-2541-2		09:31/23:42	05:27/10:38	05:32/11:43	06:07/14:25
2-5141-4		06:04/22:38	03:24/09:08	00:49/06:28	03:31/12:49
2-5351-4		04:10/19:01	05:12/09:01	04:29/08:07	05:25/13:33
2-5151-4		14:22/44:08	05:39/12:54	03:17/15:26	05:53/37:41

2-2441-2	08:19/25:04	06:40/23:08	11:17/18:15	05:50/26:17
2-5336-4	11:19/17:59	12:31/18:17	05:20/12:46	04:35/11:47
2-5353-4	13:35/25:22	04:25/16:45	06:13/14:09	05:55/25:46
2-5841-4	14:04/25:54	05:16/17:02	01:47/05:08	03:36/17:18
2-6351-4	17:02/40:00	09:30/26:06	08:22/24:30	07:05/22:01
2-1351-1	13:36/30:42	13:07/38:19	24:19/47:37	Missing
2-6331-4	08:58/30:40	05:31/17:00	05:13/25:46	05:06/25:11
2-3151-3	08:22/29:21	01:31/03:17	10:37/36:28	07:14/18:13
2-3531-3	08:03/20:01	04:52/14:13	02:44/37:48	02:56/17:05
2-5332-4	16:22/35:06	09:13/14:04	10:41/16:24	10:18/17:32
2-5343-4	06:24/17:54	10:10/21:50	02:08/11:41	4:00/13:26
2-5354-4	13:58/23:58	06:05/16:28	09:09/20:39	00:46/03:19
2-2431-2	15:05/50:00	05:58/10:59	12:58/40:09	04:43/20:52
2-3933-3	14:55/25:32	17:45/34:24	14:14/31:33	05:04/15:32
2-5333-4	10:21/20:37	07:48/18:56	01:54/07:47	09:34/21:09
2-1541-1	04:53/18:56	07:04/22:48	12:56/26:43	07:55/24:27
2-4141-3	11:25/24:58	16:22/22:48	07:50/21:17	25:08/41:37
2-4131-3	15:15/34:47	14:03/38:36	13:17/33:23	Missing
2-1531-1	09:28/25:28	00:39/02:25	05:42/18:03	Missing
Average percentage	42.8	43.36	36.60	33.20

Time measured in seconds and averaged time converted into percentage at each time point

Table C. 4. The implementation level scores of the SWH approach

Code (Treatment)	Score	Code (Control)	Score
1-0141-1	1	2-2231-2	0
1-6041-4	1.13	2-5131-4	0
1-6052-4	0.63	2-5334-4	0
1-6241-4	0.8	2-5831-4	0
1-0951-1	1.25	2-5942-4	0.56
1-0952-1	1.38	2-1451-1	0.3
1-5631-4	1.95	2-2531-2	0
1-0143-5	1.19	2-2541-2	0
1-5751-4	1.36	2-5141-4	0.5
1-6045-4	1.5	2-5351-4	0
1-6050-4	1.63	2-5151-4	0.8
1-6131-4	1.4	2-2441-2	0
1-0261-1	1.06	2-5336-4	0
1-0732-1	1.1	2-5353-4	0
1-2732-2	0.88	2-5841-4	0.25
1-0231-1	2	2-6351-4	0
1-0741-1	1.13	2-1351-1	0
1-0953-1	1	2-6331-4	0
1-3631-3	1.25	2-3151-3	0
1-3641-3	0.88	2-3531-3	0
1-5643-4	1	2-5332-4	0

1-5761-4	1.4	2-5343-4	0
1-6039-4	1.13	2-5354-4	0
1-6046-4	1.2	2-2431-2	0
1-6051-4	1.38	2-3933-3	0
1-6057-4	0.88	2-5333-4	0
1-6142-4	0.81	2-1541-1	0.13
1-2751-2	0.5	2-4141-3	0
1-6031-4	0.63	2-4131-3	0
1-6032-4	0.88	2-1531-1	0
Average	1.14	Average	0.09

The SWH approach implementation scores of the teachers in both groups were averaged across the duration of the project by adding the scores at each time point and dividing the total of the scores by the number of time points. A perfect score was 3.

Table C. 5. Topics taught in the treatment and control groups

Physics	Biology	Chemistry	Earth science
Sounds	Eco system	States of matter	Soil/rocks
Electricity	Human body	Properties of substances	Earthquakes
Motion	Classification		Volcanoes
Energy	Adaptation		Phases of the moon
Density	Life cycle of animals		Solar system
	Plants		

REFERENCES

Aguiar, O. G., Mortimer, E. F., & Scott, P. (2009). Learning from and responding to students' questions: The authoritative and dialogic tension. *Journal of Research in Science Teaching, 47*, 174–193.

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*(14), 1745-1765.

Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation: Advances in Research and Theory, 2*, 89–195.

Baird, J.R., & Northfield, J.R. (Eds.). (1992). Learning from the PEEL experience. Melbourne, Australia: Monash University Printing.

Barnes, D (1976). From communication to curriculum. Penguin Books, London.

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.

Berland, L. K., & Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, *93*(1), 26-55.

Blosser, P. E. (1973). *Handbook of effective questioning techniques*. Worthington, OH: Education Associates.

Bodrova, E., & Leong, D. J. (1998). Scaffolding emergent writing in the zone of proximal development. *Literacy Teaching and Learning, 3*(2), 1–18.

Bransford, J, Brown, AL, Cocking, RR (1999). How people learn. National Academy Press, Washington, DC.

Britton, J Talking. In: Pradl, GM eds. (1982). Prospect and retrospect: Selected essays of James Britton. Boynton/Cook, Montclair, NJ, pp. 112-122.

Bruner, J. S. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.

Bruner, J (1986). Actual minds, possible worlds. Harvard University Press, Cambridge, MA.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Carlson Powell, J., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins and effectiveness. *Colorado Springs, CO: BSCS*.

Candela, L. L. (1999). Problem based learning versus lecture: Effects on multiple choice test scores in associate degree nursing students. *Dissertation Abstracts International*, 60(5-A), 1419.

Carlsen, W.S. (1997). Never ask a question if you don't know the answer: the tension in teaching between modelling scientific argument and maintaining law and order. *Journal of Classroom Interaction* 32, 14-23

Carlsen, W. S. (1998). Engineering Design in the Classroom: Is It Good Science Education or is It Revolting? *Research in Science Education*, 28 (1), 51-63.

Carlsen, W.S. (1991). Questioning in classrooms: A sociolinguistic perspective. *Review of Educational Research*, 61, 157–178.

Carlsen, W. S. (1993). Teacher knowledge and discourse control: Quantitative evidence from novice biology teachers' classrooms. *Journal of Research in Science Teaching, 30*(5), 471-481.

Cazden, C. (1988). Classroom discourse. *Portsmouth*, NH: Heinemann.

Cazden, C. B. (2001). The language of teaching and learning. *The language of teaching and learning*.

Chin, C. (2007). Teacher Questioning in Science Classrooms: Approaches that Stimulate Productive Thinking. *Journal of Research In Science Teaching*, 44(6), 815-843.

Chin, C., & Osborne, J. (2008). Students' questions: a potential resource for teaching and learning science. *Studies in Science Education*, *44*(1), 1-39.

Clark, C. M., Marx, R. W., Stayrook, N. G., Gage, N. L., Peterson, P. L., & Winne, P. H. (1979). A factorial experiment on teacher structuring, soliciting, and reacting. *Journal of Educational Psychology*, *71*(4), 534.

Conezio, K, French, L (2002). Science in the preschool classroom. Young Children, 57, 12-18

Corson, D (1988). Oral language across the curriculum. Multilingual Matters, Clevedon.

Crawford, B. A. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching, 37,* 916-937.

Dantonio, M., & Paradise, L.V. (1988). Teacher question-answer strategy and the cognitive correspondence between teacher questions and learner responses. *Journal of Research and Development in Education*, 21, 71–76.

Davis, E. A., & Smithey, J. (2009). Beginning Teachers Moving toward Effective Elementary Science Teaching. *Science Education*, 93(4), 745-770.

Dean, D. (1986). Questioning techniques for teachers: A closer look at the process. *Contemporary Education* 57: 184-85.De Rivera, C, Girolametto, L, Greenberg, J, Weitzman, E (2005). Children's responses to educators' questions in day care play groups. *American Journal of Speech-Language Pathology*, 14, 14-26.

Dillon, J.T. (1988a). Questioning and teaching. London: Croom Helm.

Dillon, J.T. (1988b). The remedial status of student questioning. *Journal of Curriculum Studies*, 20, 197–210.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, *84*(3), 287-312.

Downing, J. E., & Gifford, V. (1996). An investigation of preservice teachers' science process skills and questioning strategies used during a demonstration science discovery lesson. *Journal of Elementary Science Education*, *8*(1), 64-75.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). Taking science to school. *Learning and teaching science in grades K-8. Washington, DC: National Academies Press*.

Erduran, S., & Osborne, J. (2005). Developing arguments. *Analyzing exemplary science teaching*, 106-115.

Everitt, B. S., Dunn, G. (1991) Applied Multivariate Data Analysis. Halsted Press, New York, NY.

Fahy, P. J. (2004). Chapter 6: Media characteristics and online learning technology. In T. Anderson & F. Elloumi (Ed.). *Theory and Practice of Online Learning*, (pp. 137- 171). Printed at Athabasca University.

Fulp, S. L. (2002). The 2000 national survey of science and mathematics education: Status of elementary school science teaching. *Chapel Hill, NC: Horizon Research, Inc*.

Gall, MD (1984). Synthesis of research on teachers. *Educational Leadership.* 42, 40-47.

Galton, M, Hargreaves, L, Wall, D, Comber, C (1999). Inside the primary classroom: 20 years on. Routledge, Boston.

Gee, J.P. (1996). *Social Linguistics and Literacies: Ideology in Discourses* (2nd edn). London: Taylor and Francis.

Gee, J. P. (2001). A sociocultural perspective on early literacy development. *Handbook of early literacy research*, *1*, 30-42.

Graesser, A.C., & Person, N.K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104-137.

Hand, B, Vaughan, P, Carolyn, W (2002). Influences of writing tasks on students' answers to recall and higher-level test questions. Research in Science Education 32: pp. 19-34

Harlen, W, Qualter, A (2004). The teaching of science in primary schools. David Fulton Publishers, London

Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury.

Jim´enez-Aleixandre, M. P.,& Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), Argumentation in science education: Perspectives from classroom-based research (pp. 3 – 28). Dordrecht, The Netherlands: Springer

Kearsley, G.P. (1976). Questions and question asking in verbal discourse: A cross-disciplinary review. *Journal of Psycholinguistic Research*, 5, 355–375.

Kelly, G. J., & Brown, C. (2003). Communicative demands of learning science through technological design: Third grade students' construction of solar energy devices. *Linguistics and Education*, *13*(4), 483-532.

Kim, S., & Hand, B. (2015). An analysis of argumentation discourse patterns in elementary teachers' science classroom discussions. *Journal of Science Teacher Education*, *26*(3), 221-236.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American educational research journal*, *31*(2), 338-368.

Lee, Y., & Kinzie, M. B. (2012). Teacher question and student response with regard to cognition and language use. *Instructional Science*, *40*(6), 857-874.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648 (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1).

Long, M.H., & Sato, C.J. (1983). Classroom foreign talk discourse: Forms and functions of teachers' questions. In: H.W.Seliger & M.H.Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 268–285). Rowley, MA: Newbury House.

Lustick, D. (2010). The priority of the question: Focus questions for sustained reasoning in science. *Journal of Science Teacher Education*, *21*(5), 495-511.

Maloch, B. (2002). Scaffolding student talk: One teacher's role in literature discussion groups. *Reading Research Quarterly*, *37*(1), 94-112.

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.

Massey, SL, Pence, KL, Justice, LM, Bowles, RP (2008). Educators' use of cognitively challenging questions in economically disadvantaged preschool classroom contexts. *Early Education and Development*, 19, 340-360.

Medley, D. M. (1977). Teacher Competence and Teacher Effectiveness. A Review of Process-Product Research.

Mehan, H. (1979). Learning lessons. Cambridge, MA: *Harvard University Press*.

McNeill, K. L., & Pimentel, D. S. (2009, April). Scientific discourse in three urban classrooms: The role of the teacher in engaging students in argumentation. In *Annual Meeting of the American Educational Research Association, San Diego, CA*.

McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education, 94*(2), 203-229.

Mills, S. R., Rice, C. T., Berliner, D. C., & Rosseau, E. W. (1980). The correspondence between teacher questions and student answers in classroom discourse. *The Journal of Experimental Educational, 48,*194–204.

Morge, M. (2005). Collective decision-making process to compose divergent interests and perspectives. *Artificial Intelligence and Law*, *13*(1), 75-92.

Mortimer, E.F., & Scott, P.H. (2003). Meaning making in secondary science classrooms. Maidenhead, *UK: Open University Press*.

National Research Council. (1996). National Science Education Standards. Washington, DC: *National Academy Press*.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, D.C., National Academy Press.

National Research Council. (2008). *Ready, set, science: Putting research to work in K-8 science classrooms*. Washington, D.C.: National Academies Press.

National Research Council. (2012). A framework for K-12 science education: Practices, cross-cutting concepts, and core ideas. Washington, *DC: National Academies Press*.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, *21*(5), 553-576.

NGSS Lead States. (2013). Next Generation Science Standards: For States, By States.

O'Connor, M.C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly*, 24, 318–335.

Ogborn, J., & Kress, G. Martins. I., & McGillicuddy, K.(1996). *Explaining science in the classroom*.

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching*, *47*(4), 422-453.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of research in science teaching*, *41*(10), 994-1020.

Pea, RD (1993). Learning scientific concepts through material and social activities: Conversational analysis meets conceptual change. Educational Psychologist 28: pp. 265-277.

Peterson, C, Jesso, B, McCabe, A (1999). Encouraging narratives in preschoolers: An intervention study. *Journal of Child Language*, 26, 49-67.

Pohl, M. (2000). Learning to Think, Thinking to Learn: Models and Strategies to Develop a Classroom Culture of Thinking. Cheltenham, Vic.: Hawker Brownlow. *See more at: http://www. nwlink. com/~ donclark/hrd/bloom. html# sthash. Iu3rhsr z. dpuf*.

Polit, D. R. (1996). *Data analysis & statistics for nursing research*. Stamford, CT: Appleton & Lange.

Pontecorvo, C. (1987). Discussing for reasoning: The role of argument in knowledge construction. *Learning and instruction: European research in an international context*, *1*, 239-250.

Prawat, RS (1993). The value of ideas: Problems versus possibilities in learning. Educational Researcher 22: pp. 5-16.

Ramsey, I., Gabbard, C., Clawson, K., Lee, L., & Henson, K. T. (1990). Questioning: An effective teaching method. *The Clearing House*, *63*(9), 420-422.

Redfield, D., & Rousseau, A. (1981). A meta-analysis of experimental research on teacher questioning behaviour. *Review of Educational Research,* 51, 237–246.

Rivard, LP, Straw, SB (2000). The effect of talk and writing on learning science: An exploratory study. Science Education 84: pp. 566-593.

Roth, W-M. (1996). Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student responses. *Journal of Research in Science Teaching*, 33, 709–736.

Rowe, M.B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37, 43–50.

Roychoudhury, A., & Roth, W. M. (1996). Interactions in an open-inquiry physics laboratory. *International Journal of Science Education*, *18*(4), 423-445.

Samson, G. K., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. *The Journal of Educational Research*, *80*(5), 290-295.

Schoenfeld, AH (1989). Ideas in the air: Speculations on small group learning, environmental and cultural influences on cognition, and epistemology. International Journal of Educational Research 13: pp. 71-88.

Schwarz, B. B., Neuman, Y., Gil, J., & Ilya, M. (2003). Construction of collective and individual knowledge in argumentation activity. *Journal of the Learning Sciences*, 12(2).

Scott, P. (1998). Teacher talk and meaning making in science classrooms: A Vygotskian analysis and review. *Studies in Science Education*, 32, 45–80.

Scott, P. (2005). Divergence or convergence? The links between teaching and research in mass higher education. *Reshaping the university. New relationships between research, scholarship and teaching*, 53-66.

Settlage, J. (1995). Children's conceptions of light in the context of a technology-based curriculum. *Science Education*, 79, 535–553.

Sinclair, J., & Coulthard, M. (1975). Towards an analysis of discourse. *London: Oxford University Press*.

Smart, J. B., & Marshall, J. C. (2013). Interactions between classroom discourse, teacher questioning, and student cognitive engagement in middle school science. *Journal of Science Teacher Education*, 24(2), 249-267.

Smith, E.L., Blakeslee, T.D., & Anderson, C.W. (1993). Teaching strategies associated with conceptual change learning in science. *Journal of Research in Science Teaching*, 20, 111–126.

Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science*, 381-405.

Sutton, G. O. (1992). Cooperative Learning Works in Mathematics. *Mathematics Teacher*, *85*(1), 63-66.

TIMSS. (2003). International mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades.

Toulmin, S. (2003). The Uses of Argument. 1958. *Cambridge: Cambridge UP*.

Van Zee, E.H., & Minstrell, J. (1997a). Reflective discourse: Developing shared understandings in a physics classroom. *International Journal of Science Education*, 19, 209–228.

Van Zee, E.H., & Minstrell, J. (1997b). Using questioning to guide student thinking. *The Journal of the Learning Sciences*, 6, 229–271

Van Zee, E.H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversations about science. *Journal of Research in Science Teaching*, 38, 159–190.

Vygotsky, LS (1978). Mind in society: The development of higher psychological processes. Harvard University Press, Cambridge.

Wajnryb, R. (1992). *Classroom observation tasks: A resource book for language teachers and trainers*. Cambridge University Press.

Walsh, JA, Sattes, BD (2005). Quality questioning: Research-based practice to engage every learner. Corwin Press, London.

Ward, B., & W. Tikunoff. (1976). The effective teacher education problem: Application of selected research results and methodology to teaching. *Journal of Teacher Education* 27: 48-52.

Waring, H.Z. (2002). Displaying substantive recipiency in seminar discussion. *Research on Language and Social Interaction*, 35, 453–479.

Wasik, BA, Bond, MA, Hindman, A (2006). The effects of a language and literacy intervention on head start children and teachers. *Journal of Educational Psychology,* 98, 63-74.

Weiss, IR, Banilower, ER, McMahon, KC, Smith, PS (2001). Report of the 2000 national survey of science and mathematics education. *Horizon Research, Inc*, Chapel Hill, NC

Weiss, I. R., & Pasley, J. D. (2004). What Is High-Quality Instruction?. *Educational Leadership,* 61(5), 24.

Whitehurst, GJ, Arnold, DS, Epstein, JN, Angell, AL, Smith, M, Fischel, JE (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, 30, 679-689.

Will, H. C. (1987). Asking good follow-up questions (Junior Great Books program). *Gifted Child Today,* 10, 32-34.

Winne, P.H. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research,* 49, 13–50.

Woolfolk, A. E., & McCune-Nicolich, L. (1984). Educational Psychology for Teachers (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Wragg, EC, Brown, G (2001). Questioning in the primary school. Routledge Falmer, London.

Zembal-Saul, C., McNeill, K. L., & Hershberger, K. (2013). *What's Your Evidence?: Engaging K-5 Students in Constructing Explanations in Science*. Pearson Education.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*. 39(1), 35 – 62.