


ORIGINAL RESEARCH ARTICLE

Open Access



Promoting productive argumentation through students' questions

May Poh Eng Phua^{1*} and Aik-Ling Tan² 

* Correspondence: phua_poh_eng@mgs.sch.edu.sg

¹Methodist Girls' School, 11, Blackmore Drive, Singapore 599986, Singapore

Full list of author information is available at the end of the article

Abstract

Questions are important in facilitating the thinking process that leads to learning. There are many research studies examining the use of students' questions as scaffolds to facilitate argument construction but more needs to be done to understand how these questions are used in generating productive arguments. As such, in this research, we investigate (1) the types of students' questions generated within a group and how these questions are used in generating productive arguments and (2) strategies used by groups of students who are deemed more successful in generating convincing arguments. Adopting a social constructivist perspective, we examined students' talk about science *within* their groups and *between* groups. We worked with a group of 24 secondary three Biology students to complete a total of seven days of crime scene investigation tasks that required them to make evidence-based decisions to determine the cause of death and solve the crime. The data collected and analyzed included transcripts from students' oral discourse and written artefacts. We found that asking hypothetical questions promotes the construction of quality arguments. Groups that were more successful in generating quality arguments adopted strategies such as using visible schema constructed from their own questions, testing the strengths of their claims and choosing claims that have the highest number of propositions.

Keywords: Argumentation, Secondary school science, Biology

Introduction

This study examined using argumentation as a strategy to enable more meaningful learning in science, particularly in enabling students to learn how to evaluate evidence and justify claims. There is rich research literature on argumentation and questioning that focused on students' questions in producing quality arguments (for example Chin & Osborne, 2010a; McNeill, 2011). However, the notion of quality varied from one study to another and to date, there has been little attempt to compare between groups of learners that were the most and the least successful in producing quality arguments. It was commonly observed that students were often weak in evaluating and justifying their claims (Osborne, Erduran, & Simon, 2004). Students might agree to a claim but often without much explanations and elaborations. Alternatively, they might disagree but fail to provide convincing reasons to support their disagreement. Counter proposing alternative solution(s) that were backed with justified reasons were also uncommon in classroom discussions amongst students. Hence, classroom discussion activities might be deemed unproductive and contribute little to bring about meaningful

learning. It is hence important to understand how questions raised by different individuals in class are linked to the social domains in learning science and if there were strategies that teachers could adopt in classroom discussion activities to promote more productive discussions and persuasive discourses amongst students. It was on these bases that engendered us to examine how students' questioning, rebuttals and counter-arguments could be orchestrated for productive argumentation to take place.

There had been a growing interest and diversity in the research of argumentation over the last few decades. We have knowledge in various aspects of argumentation, including the structure of argumentation (e.g., Sampson & Clark, 2008), designing learning environments such as using scaffolds (e.g., Chin & Osborne, 2010b; Jimenez-Aleixandre & Pereiror-Munoz, 2005; Kelly & Takao, 2002) and questions as supports (e.g., Chin, 2006), epistemology (e.g., Sandoval & Millwood, 2007), assessment of argumentation (Sampson & Clark, 2008), social aspects of argumentation (e.g., Kolsto, 2006; Mercer, 2000) and science teacher education and professional development in argumentation (e.g., Zohar, 2007). The findings in these research areas were fundamental but crucial for understanding the complexities of argumentation. However, these earlier research revealed little about how questions and ideas initiated by students could be harnessed to build nascent forms of arguments. Students bring with them a rich reservoir of prior knowledge and experiences to class and these ideas could potentially be used on their sense making interaction with their peers.

The relationship between argumentation and learning science as well as the state of argumentation in science education prompted us to ask how students made use of questions that they generated to construct sound and strong arguments in science. This study, therefore, sought to compare the differences in students' questions, types of knowledge and types of reasoning that favored the construction of a sound and strong argument. Specifically, we wanted to find out (a) the types of students' questions generated within a group and how these questions were used in generating productive arguments and (b) the strategies used by groups of students who were more successful in generating persuasive arguments.

The research questions that guided this research and their respective rationales were:

1. What questions are generated by students during group discussions and how are these questions used in constructing productive arguments?
2. What are the strategies employed by groups who are more successful in producing sound and strong arguments?

Previous research by Chin and Osborne (2010a) and Harper, Etkin, and Lin (2003) suggested that high quality argumentation was characterized by better conceptual achievement in science and was associated with the number and more importantly, the types of questions asked by students (namely, key inquiry; basic information; unknown or missing information; conditions under which the phenomena was taking place; and others). Based on their findings, we hypothesized that groups that generated a greater number of questions, in particular key inquiry questions, would be better at producing arguments of better quality (characterized by soundness and strength). We also aimed to characterize the different types of questions raised.

Earlier research studies by researchers such as Sampson and Clark (2008) also shared the significance of providing conceptual scaffolds and Chin (2006) showed how question prompts in the form of question webs could be used to help students to construct arguments. However, these earlier research had also reported groups that were less successful in generating quality arguments despite the provision of similar scaffolds and supports. In other words, using students' questions as scaffolds might not necessary enhance the development of quality arguments. As such, it might not be the provision of question prompts and webs that facilitated the generation of quality arguments but rather how students used these scaffolds to support their argument construction that were important. Hence, we hypothesized that groups that were more successful in generating quality arguments were able to devise their own aids or perhaps modify or integrate the various scaffolds and supports that were given to them to help them construct their argument.

Review of literature

Research carried out by Jimenez-Alexandre, Rodrigues, and Duschl, (2000) showed that students often found it challenging and demanding when constructing argument on their own although their argumentation skills could be improved through various forms of intervention. Examples of these interventions included teacher role modelling argumentation (Jimenez-Alexandre & Pereiror-Munoz, 2005), explicit teaching of argumentation skills (Chin & Osborne, 2010b) and the provision of conceptual scaffolds (e.g., Kelly & Takao, 2002; Sampson & Clark, 2008). In addition, there were also extensive research on the importance of extended time, role of teachers' support, students taking ownership of their learning and knowledge claims (e.g., Jimenez-Alexandre & Lopez, 2001), students reflecting about their own understanding and change in ideas, beliefs and positions during argumentation (e.g. Jimenez-Alexandre & Pereiror-Munoz, 2005). To date, the knowledge base of how to organize and structure group discussions was widespread (see Webb, 2009, for an extensive review).

In terms of assessing the quality of argumentation, many research studies had also been conducted and the work by Sampson & Clark, 2008 was of most significance. They proposed a new framework for assessing quality of argumentation after they had reviewed and identified limitations in several frameworks based on works by Erduran et al. (2004), Kelly and Takao (2002), Lawson (2003), Sandoval and Millwood (2005), and Zohar and Nemet (2001). This framework proposed by Sampson and Clark (2006) had five criteria for assessing the quality of scientific arguments. These five criteria examined the nature and quality of knowledge claim, how a claim was justified, if a claim accounted for all available evidence, how arguments attempted to discount alternatives and lastly, how epistemological reference were used to coordinate claims and evidence. Duschl (2007) conducted a study by adopting Walton's (1996) argumentation schemes which was a framework that could address most of the five criteria put forth by Sampson and Clark (2008). Duschl (2007) concluded that the use of Walton's (1996) framework offered a more productive avenue for researchers to examine quality of argumentation.

Knowledge and reasoning in argumentation

A scientifically supported claim might be built on a foundation of unsound understanding. It was not uncommon to find students providing claim(s) that were true but their

premises and/or reasoning were flawed. In other words, a student's misconceptions and unsound reasoning might be masked by a scientifically supported claim. Conversely, it was possible that the claim(s) put forth to be scientifically unsupported but the reasoning patterns were logical. For example, if I believed that all dolphins are mammals and that all mammals were fish, then it would also make sense for me to believe that dolphins are fish. Even someone who disagreed with my understanding of biological taxonomy could appreciate the consistent and reasonable way in which I used my mistaken beliefs as a foundation upon which to establish a new one. Clearly, while certain patterns of thinking and reasoning did invariably lead from truth to truth, other patterns did not.

According to Mason (1996), in the last two decades, research on learning and instruction pointed to empiricism as the basis of knowledge acquisition, that is, individuals constructed knowledge through experience. Furthermore, Pfdunt and Duit (1994) mentioned that such knowledge was often incompatible with scientific knowledge taught in schools. As a result, classroom learning required the reorganization of existing knowledge structures, that is, there was a need to design classroom learning activities to bring about conceptual change. These conceptual changes were more likely to occur when learners were required to explain, articulate, justify and evaluate during a collaborative process of knowledge construction. From this it could be seen that there was much value to determine how new knowledge was socially constructed during classroom discourse, in particular when learners came with an array of prior knowledge and experiences. In other words, it was relevant to identify the specific cognitive procedures especially the types of knowledge and reasoning used by learners during knowledge construction. More importantly, since reasons were connected and were building blocks of an argument, it would be of great significance to find out how knowledge was weaved together when learners engaged in reasoning.

This study proposed how collective argumentation could be developed through students' questions, cognitive procedures (reasoning and knowledge) as well as argument constructions. By studying the types of knowledge and collaborative reasoning embedded in argumentative activity, we illustrated how learners attempted to enlarge, fine-tune and revise their own beliefs and conceptions with their peers, thus sharing the cognitive burden of learning (Vygotsky, 1978).

Role of question-asking in argument construction

In argumentation, we hypothesized that the type of questions raised by students revealed their cognitive processes, making it a tool for assessment as well as knowledge construction. From a social-cognitive perspective, questioning amongst peers helped learners to co-construct knowledge, that is, social construction of knowledge which consequently fostered productive discussion (Chin, Brown, & Bruce, 2002). Questions articulated in the public spaces served as a means for individuals to collaborate with peers in the form of discussions, clarifications and other social forms Tan, Lee, & Cheah (2017) as they entered the "zone of proximal development" (Vygotsky, 1962, 1986).

Each student brought with him or her different prior experiences, knowledge and skills that often created discrepancies and alternative viewpoints to a problem (Tan, Lee, & Cheah (2017)). Alternative viewpoints allowed for critical thinking in science as

dissonance would foster argumentation within the group. Question-generation in the form of debates became inevitable during the resolution process. It was essential for students to recognize the various connections in logical thinking so that they could identify faulty reasoning, identify reliable evidence and construct explanations before they could refute or support a hypothesis – these were essential science process skills that fostered scientific inquiry (Curriculum Planning and Development Division [CPDD], 2013). The process could be compared to collective wisdom, where problem-solving and conflict resolution leveraged on group wisdom and co-intelligence. Hence, questioning could be deemed as a valuable tool for reasoning and thinking, especially when students built upon views of others in a collaborative setting – a feature in argumentation.

Although argumentation is seen as a process that allowed for social-construction of knowledge (multi-voiced argument), this construction of knowledge could also occur through self-reasoning and thinking. Some researchers such as Jimenez-Aleixandre and Pereiro-Munoz (2005), and Mortimer and Scott (2003) deemed this form of argument as dialogic (even when a single speaker is producing it) by taking into account the listener's perspective. According to (Vygotsky's 1962, 1986) theory of verbal self-regulation, the progression from an inter-psychological plane (collaborative discourse with questions embedded in the conversation) to intra-psychological plane (individual internal dialogue with self) is the fundamental concept in his theory. Self-questioning (Chin, 2004) is a powerful linguistic tool used to achieve self-regulation which, in turn, drive the mind to look for patterns, connections, reconcile prior experience or knowledge and make meaning. Therefore, the resultant self-verbalization in the form of self-questioning facilitates the construction of knowledge, development of concepts and metacognition through self-directed learning.

Indeed, there is great potential in argumentation as a pedagogical approach. The emphasis on quality argumentation resides at the heart of scientific literacy where higher-order thinking skills are employed to solve problems and make decisions (Hurd, 1998). If there were a way to track a student's cognitive process during argumentation, it would then be possible for us to trace their alternative conceptions in science. This would require an analysis of the students' questions and interaction patterns in the argumentation discourse. However, despite the capacity of students' questions for augmenting knowledge acquisition, amalgamation, construction and expression, much of this potential has remained untapped. Notable exceptions were the work of Chin and Osborne (2008) who found that asking critical questions during argumentation could interrogate the implicit premises of an argument, point to exceptional situations or other possible arguments. They also highlighted that questioning could serve as heuristic devices to stimulate dialectical thinking during argumentation; act as starting points for expressions of doubts, rebuttals and counter-arguments. Chin and Osborne (2010a) conducted another study that showed that questionings prompted students to articulate their puzzlement, make explicit their claims and (mis)conceptions, identify and relate relevant key concepts, construct explanations and consider alternatives especially when their ideas were challenged.

Further, a substantial body of research (e.g., Davis, 2003; McNeill, Lizotte, Krajcik, & Marx, 2006, Chin & Osborne; 2010a) had shown value in encouraging students to think and ask higher-order questions during discussions. Chin and Osborne (2010a)

conducted a study that examined and compared the questions asked by four different groups of students during discussion. They identified and provided instructional strategies to promote a more productive discourse during discussion. Productive discourse is critical for our students to be enculturated into the workings of the scientific enterprise, where co-construction of knowledge and argumentations are pivotal. Chin and Osborne's (2010a) study was believed to be the first to examine and identify the characteristics of quality and productive arguments. Their construct of productive discourse was based on the variety of questions generated across a range of question types. Most of these questions addressed key inquiry concepts, basic information and the elements of arguments. However, the differences between groups that were more successful and least successful in producing sound and strong arguments were not well understood. There is also a need to understand how questions are used in schools of different social contexts to support argumentation (Chin & Osborne, 2010a). In this study, the construct of *sound* argument was one that had accurate and coherent reasoning, premise(s) and claim(s) while the construct of *strong* argument was one that contained prepositions to support a claim and refute other claims. With the potential that students' questions and question-asking could harvest in the field of argumentation, the synergy of questioning and argumentation in the learning of science could be a powerful tool that catalyzes students' own knowledge construction, learning and understanding.

Methods

Theoretical underpinnings

In this study, learning was viewed as social in nature. This perspective was based on (Vygotsky's 1962, 1986) social-cognitive ideas. Learning took place through interaction with others together with resources such as language. During the knowledge acquisition process, learners built on one another's questions, ideas, information, and experiences shared during the discourse of group work and amalgamated them to make new meanings and understandings. During group discussions, students engaged in marshalling the appropriate evidence, argued about the strength and limits of their evidence and refined their reasoning so as to make credible knowledge claims among their peers. As such, group discussion offered a rich platform for data collection in argumentation studies. Hence, the methodological perspective that was adopted in this study reflected a stance towards social constructivism.

Research design and participants

The participants comprised of 24 secondary three (grade 9) students who studied biology for their GCE 'O' Level Examinations. All 24 students were 15 years old whose academic abilities ranged from average to high based on their performances in various summative assessments during the school year. In ensuring a more homogenous academic ability amongst all groups in this study, each group comprised of two students who are rated higher ability and two who are rated average ability.

The participating students were from an all-girls' school that is an independent school located in the western part of Singapore. The students in this school were generally from families with above average to high social economic status. The school is a full school that offered education for girls from primary one to secondary four. The

biology teacher teaching this class had 12 years of teaching experience, of which nine years was spent teaching in a coeducation school and three years in the participating school. She had a Bachelor's degree in Science, majoring in Biochemistry and a Post-graduate Diploma in Education (secondary education).

In this study, for both the activities used in familiarizing argumentation and data collection, students were given a list of question prompts that was modified from the framework given in Chin (2006) for every group discussion. This list of question prompts served as a scaffolding tool to encourage students in their question-asking process. All the learning activities in this study were collaborative in nature. Within each group, after analyzing the task, each student would first work individually to brainstorm and write the questions that they had about the task. Subsequently, they got together in their assigned groups and read out their questions.

A scribe was nominated within each group to record all the questions posed and consolidated them onto a question web provided. Using the question web and the set of question prompts provided, students within their group posed questions to each other, discussed, defended and justified their claims. To further help students to structure their arguments, each group was given two argument construction worksheets that were specifically crafted for each task. They were required to state the evidence, claim(s), warrant(s), backing(s), qualifier(s) and/or rebuttal(s).

The task that students worked on was a Crime Scene Investigation (CSI). Such tasks were complex in nature. The task required students to analyze and evaluate massive amount of data and information and covered a range of topics in the biology syllabus such as cellular transport, movement of substances, human anatomy and physiology and molecular genetics. Argumentation required the instructional context to be rich enough to enable multiple perspectives and the use of evidence to reconcile these multiple perspectives. As such, the crime scene investigation scenario was presented to the students to provide opportunities for students to develop their problem-solving skills through critical thinking as well as to present their ideas through discussions with others. Different types of problems could be designed for students to support student learning (Jonassen, 2011). At one extreme were well-structured problems, that mostly presented defined concepts within a fixed scenario and a prescribed, perfect solution. At the other extreme were ill-structured problems, that relied on a range of domain knowledge, had elements of uncertainty about the information available with regard to the problem and had multiple solutions. In the context of the crime scene investigation, we designed it such that it was along the continuum from well-structured to ill-structured. It was well-structured in the sense that there was only one solution/answer but had elements of ill-structuredness in that the process of arriving at the solution required the students to make use of logical and critical thinking to make sense of the evidence available. Hence in the design of this task to support argumentation, we took into account Berland and McNeil's (2010) four vantage point that could alter the complexity of a problem: (1) the complexity of the questions, (b) the size of the data set, (c) the appropriateness of data and (d) the level of scaffolds. Table 1 showed the degree of complexity for the four sub-dimensions in the CSI activity for the secondary three students from average to high ability.

Class activities included whole-class guided discussions, tutorials, debates, teacher demonstrations, small-group hands-on tasks, and laboratory experiments carried out in

Table 1 Degree of complexity for the four sub-dimensions in the CSI activity for average to high ability secondary three students.

Sub-dimensions	Simple	Complex
Complexity of the questions <i>Degree to which the possible answers to the question are defined and their plausibility</i>	Questions is closely defined with two to three potential answers	Questions is open with multiple potential answers
	The CSI tasks in determining time of death and cause of death have been narrowed to two potential answers for each because the key focus of the tasks is to allow students to use evidence effectively to argue and support their claim. Narrowing the possible answers to two allows more time for students to find relevant evidence to support their claim and refute the other answer.	
Size of data set <i>The number of data points and data representation with which students need to locate and evaluate</i>	Data set is small	Data set is large
	The CSI task has provided students with data from coroner's report, information from crime scene, interviews from alibi and suspects. Hence, students do not need to locate for relevant data. But all these data need to be evaluated to solve the crime hence it is scored as moderately complex for size of data set	
Appropriateness of data <i>Whether there is irrelevant information included in the data set</i>	Data set is limited to appropriate data	Data set contains both appropriate and inappropriate data
	As in the real-life context, the CSI tasks have data that are crucial to help students solve the crime and there are also data that are not useful and can cloud and influence their decision making. Students therefore need to evaluate and decide which are appropriate to be use as evidence when they need to support their claim during argumentation.	
Levels of scaffolds <i>Temporary supporting structures provided by tools or individuals to support student learning of complex problem solving</i>	Detailed scaffolds	Moderate scaffolds
	During the CSI activity students are not allowed to consult their teachers. However, given that they are secondary 3 students of average ability, templates are provided to help students consolidate the evidence, externalize their questions arising from puzzlement and organize their thinking. A template of an argumentation map is also available to help them construct their arguments. Since, some form of support it given to help them solve the problem the level of scaffold is scored at a moderate level of complexity.	

pairs or individually. For activities other than lectures, students generally first worked on a given problem, either individually or in groups of fours.

The students were divided into groups of four to work on this investigation. After we briefed the students on the activities and tasks, they proceeded to conduct laboratory investigations on the specimen collected from the crime scene. Students were given various documents that were related to the crime (for example, coroner's report, transcripts of interviews with suspects) to continue processing at home. They were required to make sense of the massive amount of data and information presented to them. They were also told to list questions that they had and bring them along for the next lesson. A scribe for each group was appointed by members within each group. Each member in the group took turns to pose their questions while the scribe recorded and pooled questions using a question web. The students were given two possible claims for the time of death and cause of death and were told to decide on one for each.

Next, with the help of their question web, students asked the questions they had listed earlier to each other and answered them with the help of information in the evidence sheet and reports on the crime to justify their choice of the claim made. Following this, they were provided with an argument template to guide them in constructing their argument map on a butcher sheet. Lastly, groups presented their arguments to the whole class in the form of a written argument maps and these were displayed in their classroom. Over the next four days, the students questioned, clarified, probed or challenged each other with alternative theories, counterarguments and rebuttals using sticky notes during a gallery walk.

During the process of determining the cause and time of death, members in the group argued and justified their reasons for the elimination of the hypotheses, made assumptions and justified their assumptions with existing evidences and proposed the need for collecting new evidence from the crime scene and/or coroner. The students

were immersed in an environment where they were required to make evidence-based decisions to help them solve the crime.

Data collection

In this study, oral and written data were collected. The nature of argumentation in a social context naturally fostered students to engage in a collaborative discourse on scientific reasoning. However, several recent studies of science education provided evidence for the importance of writing in students coming to understand and use scientific concepts (Keys 1999; Keys et al. 1999; Prain & Hand, 1999; Rivard & Straw, 2000), as well as learning to participate science as a learning community (Chinn & Hilgers, 2000). These educational studies recognized that writing and argument played important roles in scientists' thinking and reasoning. In addition, according to Kelly et al. (2000), students wrote not only to master the concepts but also to develop competencies in the specifics of argumentative practices.

As students engaged in serious writing, they moved beyond a simple formal approach for science to active work with scientific evidence, knowledge and concepts, thereby developing their thinking, reasoning and communicative skills essential for learning and writing science. This allowed researchers to analyze the cognition and metacognition of students expressed in a written form. Hence, oral and written data were collected to capture a more enriched data pool for analysis.

Data analysis

As the amount of data was vast, the research questions guided the data analysis. The procedures for the analyses were summarized in Table 2.

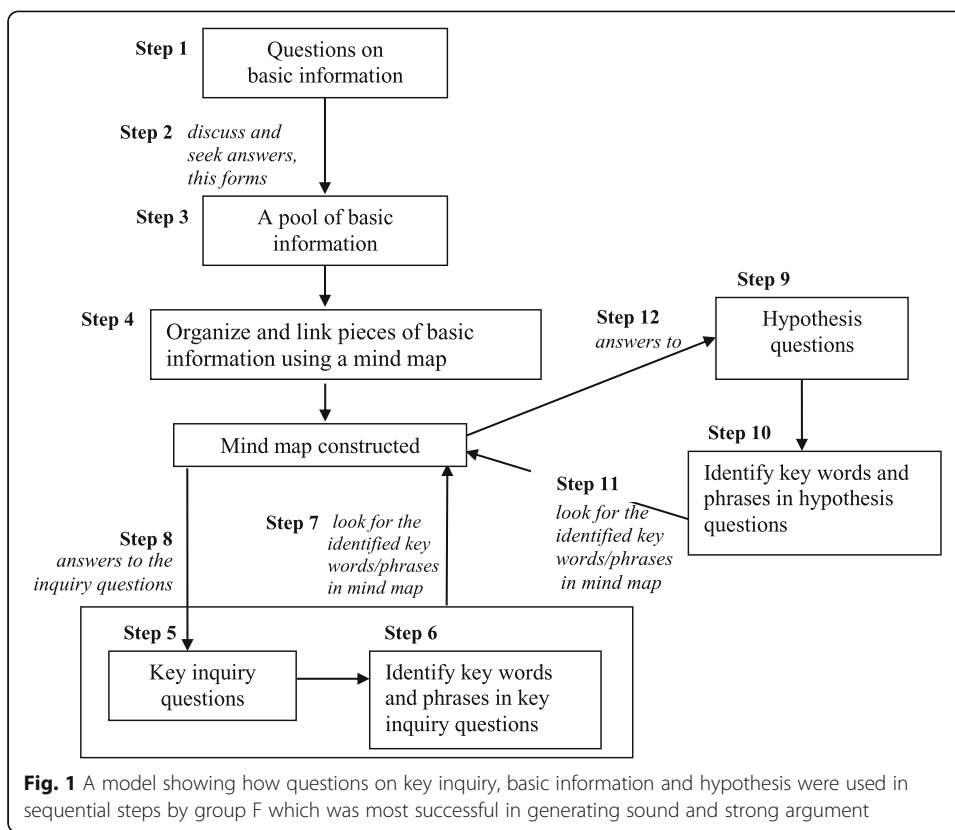
The categories of questions were similar to those from the study by Chin and Osborne (2010a) with the exception of questions on the conditions of the experiment. Given the investigative nature of the tasks and activities used in the crime scene investigation, students were likely to build hypotheses and find evidence to verify or eliminate these hypotheses before coming to a verdict. Hence, condition-typed questions were not found. Instead, a new category of question that involved hypothesis and predictive thinking emerged. Consequently, the categories of questions for this study were referred and analysed according to Chin and Osborne's (2010a) categories of questions which included: (a) key inquiry, (b) basic information, (c) unknown/missing information, (d) hypothesis and (e) others. Questions on hypothesis referred to questions that showed consideration of alternatives by showing attempts to make educated guess or informed prediction of possibilities and not just limited to changing variables or conditions (Lawson, 2003). The answers to the questions now formed a pool of basic information in step 3 Fig. 1. In step 4, the pieces of basic information gathered were linked together on the basis of cause-and-effect and effect-and-cause relationships, forming a mind map. In steps 5 and 6, key inquiry questions were asked, key words and phrases were identified from the key inquiry questions. In step 7, students then looked for these key words and phrases in the mind map and attempted to identify them. Once located, they then traced and checked back the various links in the map. With these links, students weaved them into coherent pieces of propositions that addressed the key inquiry questions and this constituted step 8. Finally, in steps 9 to 12, to test the strength of

Table 2 A summary of data analysis for each research question

Research question 1: What questions are generated by students during group discussions and how are these questions used in constructing productive arguments?	
Step 1	Identified all questions written artefacts and transcripts.
Step 2	Grouped all questions into different categories: a) Key inquiry questions b) Basic information questions c) Questions on unknown/missing information d) Hypothetical questions e) Other While grouping, looked out for emerging category of questions.
Step 3	Determined the number of questions generated, number of different types of questions generated.
Step 4	Scored each map against an argument scoring scheme to determine soundness and strength.
Step 5	Used SPSS and determine the correlation between: a) number of questions generated and soundness of argument b) number of questions generated and strength of argument c) number of different types of questions and soundness of argument d) number of different types of questions and strength of argument
Step 6	Identified the types of questions that showed the highest correlation to strength and soundness.
Step 7	Identified the groups who are the most and the least successful in generating quality argument based on the scores obtained for soundness and strength.
Step 8	Examined the transcript of the verbal discourse from the groups identified in step 7.
Step 9	Segmented the transcript from each group into episodes.
Step 10	Analyzed all episodes to determine how the questions (questions here refers to the types of questions identified in step 6) and speech moves were related to one another, how they were organized and structured towards a strategy to help students support or refute a claim.
Research question 2: What are the strategies employed by groups who are more successful in producing sound and strong arguments?	
Step 1	Identified the groups who are more successful in generating quality argument based on the scores obtained for soundness and strength.
Step 2	Examined the transcript of the verbal discourse from the groups identified in step 1
Step 3	Segmented the transcript from each group into episodes.
Step 4	Analyzed all episodes to identify the features of discursive interactions that led to the construction of quality arguments.

their arguments, they asked hypothetical questions and made informed predictions with the help of the mind map in a similar approach used in steps 5 to 8. In doing so, the members in the group underwent an elaborated and structured cognitive exercise with the help of a visible schema – the mind map. In Fig. 1, our model illustrated students beginning their discourse in step 1 by listing questions that sought basic information followed by step 2 where they answered the questions that they have listed.

The number of questions and the number of different types of questions generated were correlated with the scoring for *soundness* and *strength*, using SPSS. The statistical data obtained were used to determine the correlation between the total number of questions and each type of question to the *soundness* and *strength* of the argument. Finding the correlation between total number and number of types of question asked to that of *soundness* and *strengths* was important because if the correlation was a strong one, we could then be assured that there was a positive relationship between total number of questions asked to the quality of argument (measured based on soundness and strengths) and also a positive relationship between types of questions asked and the quality of argument (measured based on soundness and strengths). In essence, RQ1 and RQ2 were linked. If a positive relationship of the above could be established,



we could confidently put forward the claim that the type and numbers of question types posed by the most successful group did indeed facilitated the group to generate quality arguments (since there was a positive relationship between total number of questions, types of questions to that of soundness and strengths). Subsequently, we proceeded to examine how these questions from the most successful and least successful group were used in constructing arguments. From there, we created a model of how questions were weaved together during construction of quality arguments (see Fig. 1). Figure 1 was developed based on our observations and subsequent inferences on how the groups worked together.

For our analysis, although a scheme was provided, we were mindful that the activity used by Chin and Osborne (2010a) was different from this present study, hence not all the students’ questions in this study would fit into the scheme from Chin and Osborne (2010a) and that new types of question might emerge. The categorizing of questions was done independently by the first author and another science educator who had eight years of teaching experience in secondary schools and had no stake in this study. The inter-rater reliability of scoring was 90% after discussion. After categorizing, the questions in each category for each group were counted. In addition, the types of concepts addressed by each category of questions were identified and tallied.

The analysis of the *soundness* of the arguments was based on the truth and falsity of their premises as well as the validity of the inferences. In other words, it required students to employ both reasoning skills as well as questioning skills. To determine the quality of the written argument, each argument map was scored against an argument scoring scheme. The *soundness* of the argument was determined by the total number

of propositions comprising of true premises with valid inference. One point was awarded for a set of propositions made in each warrant-backing chain that contained supported premises and valid inferences. For example, in one group, under warrant: *An air bubble may block the coronary artery (premise) leading to muscle death (inference)* and under backing: *Air bubbles travel up radial vein into the left ventricle through the hole in septic defect and into coronary artery (premise), blocked the blood flow, no oxygen to support respiration, leading to muscle death (inference)*. This warrant-backing chain constituted a proposition set and was awarded one point.

The *strength* of the argument was determined by the total number of proposition sets and this was based on the scoring scheme shown in Table 3. Each proposition set might support the scientifically valid claim (the scientifically valid claim was intravenous injection) or refuted the scientifically invalid claim (the scientifically invalid claim was drowning). Table 3 illustrated all the proposition sets that could possibly be used during argument construction for this particular activity. The crime scene investigation and its storyline was designed to include only the propositions found in Table 3 as evidence for the time and cause of death. All the propositions in Table 3 were the only possible ones that have sound premise and valid inference. While certain patterns of thinking and reasoning did lead from truth to truth, other patterns did not. It was possible for a student or group to correctly identify the cause of death but the reasoning was flawed. Therefore, while it was possible for students to come out with other proposition sets, the proposition sets other than those found in Table 3 were reasoning and/or premises that were flawed. In terms of scoring the map, one point was given for each proposition set. The maximum possible score by any group for strength was five since there could only be five possible proposition sets. The scores for *soundness* and *strength* were then further processed to obtain their percentage.

Results and discussions

Relationships between questions, soundness and strength of arguments

The distribution of questions by numbers and types for all six groups were examined. Each group wrote an average of 16 questions and asked an average of 24 questions. Most questions were key inquires (42.6%), basic information (21.1%) and hypothesis (23.1%). The questions raised by the students from various groups were conceptually different. There were (10 concepts) used to account for the cause of death. Table 4 showed a distribution of these questions and Table 5 showed examples of the types of questions asked by students. Since, most of the questions asked were key inquires, basic information and hypothetical ones, it followed that the correlation between these questions to the *soundness* and *strength* of the arguments constructed should be examined to determine if the type of questions raised contributed to the *soundness* and *strength* of the argument. Table 6 hence showed the correlation analysis of key inquiry questions, questions that sought basic information, hypothetical questions and total number of questions asked to *soundness* and *strength*.

There was a high, positive and significant correlation between the following questions and the *strength* of argument constructed: a) total number of questions ($r = 0.92$, $p = 0.01$), b) number of key inquiry questions ($r = 0.77$, $p = 0.06$) and c) number of hypothetical questions asked ($r = 0.78$, $p = 0.06$). However, the correlations of these

Table 3 Scoring scheme for strength of argument

Component of argument	Description	Score
Determining the cause of death: Support intravenous injection		
Evidence 1	Livor Mortis showed dark purplish red distribution found on back of shoulders, back of arms, back of waist and back of legs	1
Warrant 1	Victim was lying on her back when she died – those area that does not comes in contact with her bed showed no discolouration	
Backing 1	Blood stopped circulating when the heart stopped beating. Blood cells settled (due to gravity) to the lowest areas in the body and accumulates in areas not in contact with any surface since the blood capillaries were not compressed.	
Evidence 2	-Vial containing synthetic hormones (estrogen and anti-androgen) -Fresh jabs and healing wounds -Blood stain on needle Injected left and right radial vein showing fresh and old injection marks	1
Warrant 2	Hormones being injected in the deceased body as needed hormone to remain woman	
Backing 2	Victim has XY chromosomes indicating that victim is a male by birth. Estrogen is a female hormone. Victim has intravenous injection for hormone therapy on regular basis	
Evidence 3	Blockage in coronary artery leading to muscle death	1
Warrant 3	An air bubble/ air embolism may have blocked the coronary artery leading to muscle death	
Backing 3	-A correct description of how air embolism can lead to restriction of blood flow to coronary arteries in atrial septic defective individual (this should include the blood vessels involved). -Limited supply of oxygenated blood to cardiac muscles. With the lack of oxygen and nutrients, cardiac muscles eventually die.	
Determining the cause of death: Refute drowning		
Evidence 4	-Presence of red fluid, mixture of river water (similar to those in crime scene 1) and haemoglobin can be found in the lungs -Blood extracted from pulmonary capillaries showed lysis of red blood cells	1
Warrant 4	Pressure in the victim's lungs was lower than her surrounding Osmosis occurs since river water had a higher water potential than the red blood cells.	
Backing 4	River water entered lungs due to pressure differences, water molecules moved from alveoli to pulmonary capillaries. Water molecules entered into red blood cells and lysed red blood cells, releasing haemoglobin which diffused across into the alveoli spaces.	
Evidence 5	High level of potassium ions in blood	1
Warrant 5	Lysis of red blood cells and cardiac muscle damage due to warrant 3 and backing 3 releases potassium ions.	
Backing 5	Red blood cells and cardiac muscle contains potassium ions	
Qualifier	–	
Claim	So, the cause of death was cardiac arrest due to air embolism as a result of intravenous injection	

three groups of questions to the *soundness* of arguments were low with: a) total number of questions ($r = 0.14$, $p = 0.80$), b) number of key inquiry questions ($r = -0.11$, $p = 0.84$), and c) number of hypothetical questions asked ($r = 0.19$, $p = 0.72$) (see Table 6). Key inquiry questions sought to explain or address big fundamental ideas while hypothetical questions suggested alternatives by showing attempts to make educated guesses or informed predictions of possibilities. The nature of these two types of questions allowed students to examine and address the problem or task from different perspectives. Consequently, they were able to organize their arguments to include propositions that could support their claim and refute the others. The ability to

Table 4 Distribution of oral and written questions by numbers and types for all six students groups

Group/Types of questions	Number of questions					Asked by each group
	Key inquiry	Basic information	Missing information	Hypothesis	Others	
A	16	11	3	2	6	38
B	9	9	0	13	2	33
C	29	5	5	15	4	58
D	11	8	3	5	2	29
E	12	6	4	3	1	26
F	26	12	1	18	1	58
Total number of question in each category	103	51	16	56	16	242

Note: Mean number of questions per group = 40
Standard deviation of oral and written questions = 14.26

support a claim and refute other claims was the feature of *strong* arguments. In addition, students needed to be divergent in their thinking and reasoning before converging to their choice of claim. Examining the problem from different dimensions probably explained for the high correlation for key inquiry and hypothetical questions to the *strength* of argument constructed. Most significantly, perhaps, was the finding that the correlation for hypothetical questions was similar to that of key inquiry questions. Questions that were hypothetical in nature seemed to have the potential to promote more divergence in thinking as they allowed students to explore all possibilities and options, suggested alternatives and made predications.

While seeking correct and accurate information was the foundation to a sound argument, students' ability to weave pieces of evidence together in a logical and coherent manner to justify a claim was equally important. Consequently, asking questions that sought basic information became essential for the construction of sound arguments. This explained for the relatively high, positive, and significant correlation between the number of questions that sought basic information and the *soundness* of arguments ($r = 0.79$, $p = 0.06$).

To confirm our interpretations, we examined the statistical data for a specific group (group C). From Table 4, it could be seen that although group C generated as many questions as group F, the number of questions seeking for basic information was the lowest amongst the six groups. It was likely that these basic questions were only sufficient to support the construction of sound propositions that supported a claim but insufficient to generate sound propositions to refute the others. Consequently, even though group C had generated a high percentage of hypothetical questions that explained for the strong argument to support the claim, their low number of basic questions had resulted in a low score for the degree of *soundness* (see Table 6).

Differences in questions asked between group E and F

The preceding finding revealed the intimate relationship between questions asked and the constructions of quality arguments that were *sound* and *strong*. In what follows, we analysed and compared questions generated by the two groups, one who was least successful and the other who was most successful in generating quality arguments.

Table 5 Types of questions (adapted from Chin & Osborne, 2010) asked by students (oral and written forms)

Question types and Concepts Addressed	Examples of Questions
<i>Key inquiry: Sought to explain or addresses big fundamental ideas related to the cause of death</i>	
Heart structures	What role does the septic defect had on her death? Why is cardiac muscle death only located at the left ventricle?
Bubble	How did the bubble get into the bloodstream and travel to the coronary artery?
Water in lungs	Why did water flow into the lungs?
Pressure differences	How do pressure differences explain for direction of water flow?
Basic information: Factual information students needed to know before they could answer question on key inquiry and hypothesis.	
Lysis of red blood cells	Is osmosis the cause of red blood cells lysis? What caused the red blood cells to agglutinate?
Hormones	What is the function of oestrogen? What is anti-androgen?
Gender identity	Is he born a male? What does XY implies?
Blockage of blood vessel	What could block the coronary artery?
Unknown/missing information: Any information required to determine cause of death but was not given.	
Intravenous injection	Was the injection voluntary?
Blockage of blood vessel	Is she a smoker since smoking can damage heart and coronary artery?
Hypothesis: Suggested alternatives by showing attempts to make educated guess or informed predication of possibilities	
Intravenous injection	What if someone deliberately injected bubbles into her? What if she was killed by intravenous injection and then later dumped in the river?
Water in lungs	What if she was drowned?
Potassium ions in blood	Could the high level of potassium ions be due to cardiac muscle death instead of lysis of red blood cells?
Others: Miscellaneous questions that do not fall into the above categories	
Structure of argument	Is this under warrant or backing?
Relevance	Which evidence is relevant for this?
Validity	Are you sure this reason is valid?

Note: Questions above were sampled from across all the six groups from their oral discourse and written artefacts

As mentioned earlier, the most and the least successful groups in constructing quality arguments were determined by the scores for *soundness*, *strength* and average percentage of *soundness* and *strength*. In addition, the most successful group must show proposition sets that supported and refuted claims while the absence of this feature might be observed in the least successful group. From our analysis of *strength* and *soundness*, group F scored the highest average percentage (80%) while group E scored the lowest

Table 6 Correlation coefficient of question types and total number of questions to soundness and strength of argument constructed

Types of questions	Key inquiry	Basic information	Hypothesis	Total number of question asked
correlation to:				
soundness	-0.11 (0.84)	0.79 (0.06)	0.19 (0.72)	0.14 (0.80)
strength	0.77 (0.06)	0.34 (0.51)	0.78 (0.06)	0.92 (0.01)

Note: Numbers indicate Pearson correlation coefficient. Numbers in parentheses indicate *p* value

(22.75%). Group F was more successful in producing arguments that comprised of more chains of propositions that had true premises, valid inferences and scientifically valid claim. In addition, the propositions generated by group F were more varied as their arguments comprised of propositions that supported one claim and refuted the other. Group F also matched 80% of the teacher's scoring scheme. It was important to note that although groups A and B also scored the same average percentage (80%) as group F but the propositions generated by group A only presented propositions that supported one claim. As for group B, the propositions (with sound premises and valid inferences) generated were lesser compared to group F. Consequently, group F was deemed the most successful in constructing arguments.

On the other extreme, group E was the only group amongst the six that incorrectly claimed the cause of death to be due to drowning. Although group E constructed varied propositions to support their claim of choice and refuted the other, they only managed to get 20% of their propositions to match the teacher's scoring scheme. Furthermore, out of the four chains of propositions three were flawed since the premises were not supported and/or inferences not valid. Consequently, group E was deemed the least successful in constructing a *sound* and *strong* argument. In terms of the *number* of questions generated, group F generated 55.2% more questions than group E. In terms of the *types* of questions generated, group F generated 53.9%, 50.0% and 83.3% more key inquiry, basic information and hypothetical question than group E respectively. In terms of the *science concepts* addressed during the argumentation process, group F addressed all the ten while group E only addressed four.

Group F's questioning and patterns of discursive interaction showed a more meticulous effort to first ensure that all questions that sought basic information were addressed. The students posted more specific questions pertaining to the various data in the coroner's report. For example, "What cause the red blood cell to lyse?" and "what can cause red blood cells to agglutinate?" The answers to these basic information questions were then organized using a mind map that showed how a piece of basic information could be linked to the others. Next, the students then asked key inquiry questions that were more open ended. For example, "how did the bubble get into the bloodstream and travel to the coronary artery?", "why was the lungs filled with water?", "why is the cardiac muscle death only located at the left ventricle?" and "what role does septic defect had on her death?" Using the mind map, students tried to identify key words or phrases used in the inquiry questions. Once identified, they tried to trace and backtrack all the possible paths that branched from the identified words or phrases. Lastly, they proceeded to test the strength of their reasoning by posting questions that were hypothetical and applied the reasoning of how the sequence of events unfolded to a hypothetical context. For example, "What if he was drowned?", "what if someone deliberately injected bubbles into her?" and "what if she was killed by intravenous injection and then later dumped in the river?"

Consider the following transcripts that were extracted from group F's discussion (Excerpt 1). Group F had just finished answering the questions that they have posed. In line 1 of Table 7, student C asked to confirm if all the questions seeking basic information were addressed. Lines 2 to 9 showed students in group F discussing about one last question that was not answered. This discussion was initiated by student Y who asked a question to seek basic information on the cause of red blood cells lysis (see line 2).

Table 7 Excerpt 1: A segment of discourse from group F's discussion (lines 1 to 10)

Turn of talk	Student	Discourse
1	C:	Have we cleared all questions?
2	Y:	Let me check, okay, I have one more? RBC lyse...what was the cause?
3	S:	Can be caused by agglutination or osmosis.
4	C:	Was she injected with something that caused her red blood cell to agglutinate?
5	S:	Oestrogen and anti-androgen are hormones. Don't think they will cause agglutination otherwise she would be dead long ago right? I mean she was on regular injection.
6	L:	Then, osmosis. If water entered the red blood cells, the red blood cells would become bloated and eventually burst.
7	Y:	Hmmm...for water to enter red blood cells, the external must have higher water potential than the cytoplasm of red blood cell. But plasma concentration should be isotonic to that of red blood cells.
8	L:	Not unless if water entered her lungs and moved into her blood stream, then her blood would have a higher water potential.
9	S:	Yah, that also explains for the presence of river water in the lungs...so clever!
10	C:	Let's link these together in the mind map.

With this last question answered, group F then organized all the information they had recorded from answering questions that sought basic information into a mind map (see line 10). Figure 2 showed the mind map constructed by group F.

They proceeded to seek answers to their key inquiry question (Excerpt 2). In line 11 of Table 8, student L asked two key inquiry questions: "how did the bubble get into the blood stream?" and "how did it eventually end up at the coronary artery?". The group then proceeded to look for key words in the key inquiry question asked. This was evident in line 13 of the transcript where student S identified bubbles and blocked coronary artery as the key words. Once the key words were identified, they traced and backtracked all possible paths that branched from the identified words or phrases in the mind map. This was evident when line 13 and Fig. 2 were juxtaposed. In line 13, student C pointed to the "bubbles enter the bloodstream" found at point F on the mind map. Student C said "Bubbles enter bloodstream, so it is here, trace backwards, that's from intravenous injection (refer to point D on mind map)...why need injection...because need to stay as a female

Table 8 Excerpt 2: A segment of discourse from group F's discussion (lines 11 to 14)

Turn of talk	Student	Discourse
11	L:	How did the bubble get into the blood stream? And also how did it eventually end up at the coronary artery?
12	S:	Where shall we start? "Bubbles" or "block coronary artery?"
13	C:	Let's start from bubbles.... <i>Student C pointing to bubbles enter bloodstream found on the mind map.</i> Bubbles enter bloodstream, so it is here, trace backwards, that's from intravenous injection...why need injection...because need to stay as a female ...why must stay as female...because born as male...XY in karyotype.
14	S:	Then, from bubbles in blood stream again, we move forward, travel into right atrium, then move into left atrium cross the septic defect hole...down the left ventricle...up the aorta, into coronary artery...block blood flow...no oxygen and nutrient... cannot respire... cardiac muscle death... dead muscle release K ⁺ ions ...detected high in blood plasma.

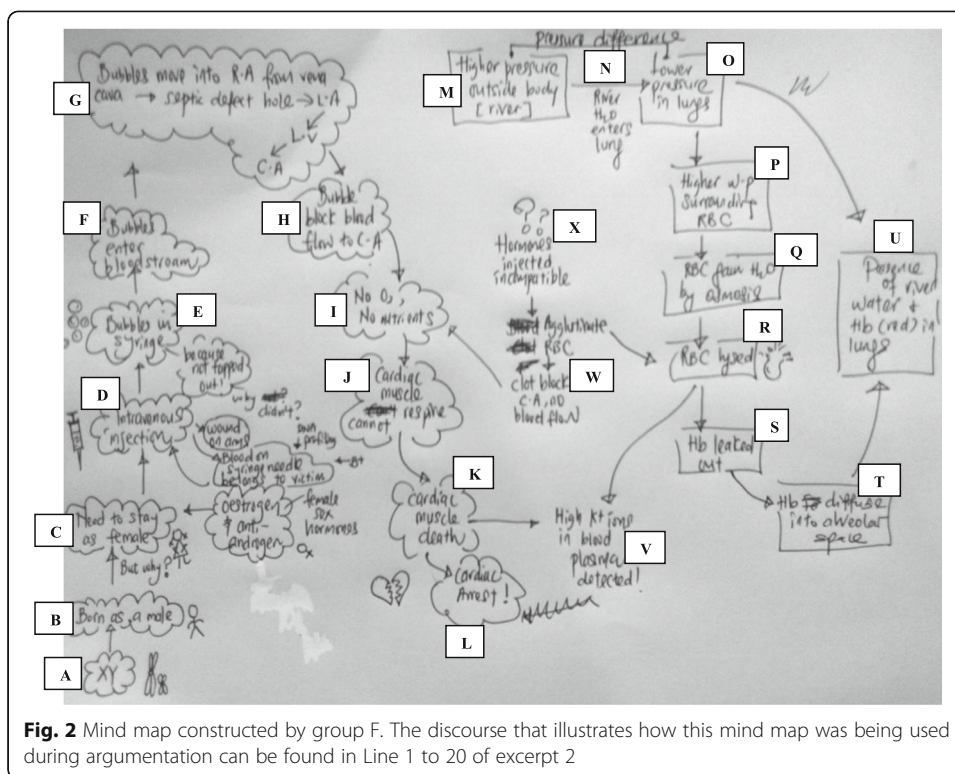


Fig. 2 Mind map constructed by group F. The discourse that illustrates how this mind map was being used during argumentation can be found in Line 1 to 20 of excerpt 2

(refer to point C on the mind map) ... because born as a male (refer to point B on the mind map)...XY in karyotype (refer to point A on mind map)”.

Finally, a hypothetical question (see line 15 of Table 9: “What if she was killed by intravenous injection then later dumped into the river?”) was posed and they went through a similar process as before to determine if these questions could be answered in a logical manner using the mind map. This was evident when we juxtaposed line 18 of Table 9 to Fig. 2. In line 18, Student Y said, “Okay for the drowning part...on our map, drowning so means lungs must be filled with water ...water in lungs...river water enter lungs (refer to point N on the mind map)...pressure differences (refer to point M

Table 9 Excerpt 3: A segment of discourse from group F’s discussion (lines 15 to 20)

Turn of talk	Student	Discourse
15	L:	What if she was killed by intravenous injection then later dumped into the river?
16	S:	Now if bubbles not tapped out...then this path forward from bubbles to cardiac muscle death and path backwards to XY on karyotype still make sense.
17	L:	Yah, agree, then dumped into river after death.
18	Y:	Okay for the drowning part...on our map, drowning so means lungs must be filled with water...water in lungs ... river water entered lungs ...pressure differences...Hmm...so it is possible!
19	L:	From water in lungs ...water diffused into bloodstream and increased water potential ... water entered RBC cells by osmosis ... RBC lysed...release K+ ions... detected high in blood plasma
20	C:	Hey...this can also lead to high k + ions in blood plasma. Maybe the high is from both sources.

and O on the mind map)...Hmmm...so it is possible!”. Hence, the turns of talk in lines 15 to 18 could be seen as an attempt to verify or eliminate the hypothesis.

It was noted that in line 20, the students realised that there could be two possible sources for the high potassium ions in blood plasma. This suggested that through answering the hypothetical question posed by student L (see line 15), members in group F were directed to think more divergently. Consequently, the hypothetical questions allowed the students to explore other perspectives and make more linkages between the information. This was evident when lines 19 and 20 were analysed together with Fig. 2. In line 19, student L said “From water in lungs ...water diffuse into bloodstream and increase water potential (refer to point P on the mind map) ...water enter RBC cells by osmosis (refer to point Q on the mind map) ... RBC lysed (refer to point R on the mind map)...release K^+ ions ... detected high in blood plasma (refer to point V on the mind map)”. Student C then realised the other possible source for high potassium ions and exclaimed in line 20, “Hey...this can also lead to high K^+ ions in blood plasma. Maybe the high is from both sources”. The discourse that emerged in lines 19 and 20 was initiated by a hypothetical question in line 15 and without the hypothetical questions, divergent thinking would be limited and their thought processes were likely to remain linear and as a result, students made fewer linkages. Consequently, they were not able to make more varied propositions to support a claim and refute the others.

Group E on the other hand, used a different strategy (see Table 10). The students in the group first asked key inquiry questions (see line 21 of Table 10: “What do you think is the cause of death?”). They then asked questions to seek basic information such as in line 24, “How did the water enter her lungs”, in line 26, “What could that red thingy be?” and “What are potassium ions?”, which were required to answer their key inquiry questions. The answers obtained were then weaved into propositions to answer the key inquiry questions (line 30).

At the beginning of their discussion, members of group E reached a decision about the cause of death very quickly (line 23: “I thought so too. Okay, that’s our claim”). They proceeded to find evidences and reasons to support their claim (line 23: “Let’s find all the evidence we can to support this”).

It was noted that in group E, although they asked key inquiry questions (line 21: “What do you think was the cause of death?” and line 24: “How did the water enter her lungs?”) and gathered evidence for both claims, they did not ask sufficient basic information questions to help them construct a valid argument. Obviously, the lack of basic information resulted in incorrect inferences and hence, the scientifically invalid claim. In line 32 of Table 10, the idea that hormones were not lethal and therefore could not cause cardiac arrest in itself bore much fallacy. It also suggested that students had not done enough ground work by asking and seeking basic information about what oestrogen and anti-androgen were, the purpose of injecting them, and whether these hormones could cause cardiac arrest and death. In addition, there were only a three hypothetical questions asked and this was insufficient to promote problem solving from multiple perspectives.

The discourse analysis of group E’s discussion also revealed their limited thinking about what could have resulted in the high concentration of potassium ions in the blood plasma. This was evident in line 30 where students commented and affirmed that the high potassium ions were from the lysed red blood cells. Group E was thus less

Table 10 Excerpt 4: A segment of discourse from group E's discussion (lines 21 to 32)

Turn of talk	Student	Discourse
21	E:	What do you think was the cause of death?
22	S:	I believe its drowning.
23	E:	I thought so too. Okay, that's our claim. Let's find all the evidence we can to support this.
24	A:	To be drowned, water must enter her lungs. How did the water enter her lungs?
25	S:	The killer pushed her head down into the river water, so as she grasped for air, she inhaled the water into her lungs.
26	A:	Then, beside river water, the content in the lungs was red? What could that red thingy be? Also, what are potassium ions? How come they are found so high in the blood plasma?
27	S:	The red thingy is haemoglobin.
28	A:	But haemoglobin is found inside the RBC.
29	S:	Yes, but RBC has lysed because they gained water by osmosis so haemoglobin released.
30	E:	RBC also contains potassium ions. So when RBC burst, these ions were released and remember Mrs. Liow mentioned when the ions are out-of-balance in our plasma, can lead to cardiac arrest. So it must be this that explains for the high potassium ions.
31	A:	Then what about the intravenous injection? Hormones ... what was that for?
32	E:	The content of the injection is hormones. Hormones are not lethal and not strong enough to cause cardiac arrest.

divergent in their thinking and this perhaps was a consequent of insufficient hypothetical question being asked. In other words, the lack of or insufficiency in hypothetical questions asked during group discussion could become obstacles for optimum and fluent cognitive operations for reasoning, problem solving and decision making.

In contrast, group F's divergent thinking and reasoning was a spin off from their attempt to hypothesize the possibility of the victim being killed by intravenous injection and later disposed of her body into the river (see line 15 of Table 9 for the hypothetical question asked). Clearly, group F's questioning and pattern of discursive interactions showed a more elaborated process of questioning to seek basic information, organizing and connecting these basic information, identifying and backtracking to find answers to their key inquiry questions. They then asked hypothetical questions to test the strength of the relationship between basic information as well as between basic information and key inquiry questions. This repeated verification and clarification of the various pieces of evidence had lead to an argument that was both *sound* and *strong*.

Due to the extensiveness of the findings on the differences between group E and F during argument construction, we have summarized the findings in Table 11.

How questions are used by students who are most successful in generating arguments that are sound and strong

The present study showed that groups that were able to generate more questions were more able to construct quality arguments. More significantly, questions on key inquiry, basic information and hypothesis when asked more frequently generated arguments that scored higher in terms of degree of soundness and strength. Similar findings were first discussed by Harper, Etkina, and Lin (2003) who found that though the number of questions asked played a significant role in the production of quality arguments, it is the types of questions that were instrumental in acquiring better conceptual

Table 11 Differences between group E and F during argument construction

Features	Group E	Group F
Number and types of questions generated by groups		
Key inquiry questions	12	26
Questions seeking for basic information	6	12
Questions seeking for missing information	4	1
Hypothetical questions	3	18
Other questions	1	1
Total number of questions	26	58
Science concepts addressed		
Number of science concepts addressed	4	7
Quality of argument constructed		
Degree of soundness/ %	25.5	80.0
Strength/ %	20.0	80.0
Average/ %	22.8	80.0
Argumentation processes		
Characteristics of discursive interactions	- Lesser deliberations and linear in thinking - Agreed on the claim (cause of death) immediately after answering questions posed.	- Elaborated, meticulous and more divergent in thinking - Agreed on the claim (cause of death) after a long discussion and deliberation.
Sequence of argumentation	1. Listed all questions they had about the problem 2. Answered all questions in random order 3. Decided and agreed on the claim (cause of death) 4. Discussed at random about evidence and reasons to support the claim. No specific patterns of discursive interaction was observed	1. Listed all questions they had about the problem 2. Answered all questions seeking for basic information first 3. Organized answers into a mind map 4. Asked key inquiry questions 5. Identified keywords in key inquiry questions 6. Looked for these keywords in the mind map 7. Traced and backtracked all possible paths that branches from the keywords 8. Asked hypothetical questions (or gave a hypothesis), look for key words in the hypothetical questions and repeat step 6 and 7 above to verify and eliminate the hypothesis
Use of visual aids	Did not use any form of visual aids	Mind map was constructed by students to organize basic information, tracked the sequence of events, and made linkages between evidences and propositions.
*Duration of productive discussion	43 min	81 min
Testing strength of arguments	No	Yes by asking hypothetical questions to explore alternatives

*Note: *duration of productive discussion refers to the amount of time taken by the group when they were on-task discussing about the CSI activity. Time spent discussing about other issues not related to the CSI (e.g., when is their mathematics homework due for submission? What are they going to have for lunch later?) is not included in the above tabulation*

achievements in physics. In the same vein, Chin and Osborne (2010a) also reported that productive argumentation was characterized by students' questions that focused on their key inquiry questions. However, what stood out from this study is the relationship between key inquiry questions, question seeking for basic information and hypothetical questions and how these questions were employed by the most and least successful groups during argument construction. Figure 1 showed a model that offered a possible representation of how these three categories of questions interacted with one another during their oral discourse.

When students ploughed through the questions on basic information and sought answers to these questions, they had opportunities to argue, clarify and verify within their group the validity of the concepts that formed a pool of propositions for the basic information. We argued that this was critical in ensuring that the propositions were sound and we based this argument on the high correlation between questions that sought basic information and soundness of argument ($r = 0.79$, $p = 0.06$). The procedure of marking out key words and phrases and locating them in the mind map was instrumental in facilitating the students to construct argument from multiple perspectives since the dynamic relationships of all basic information was made visible through the lens of the mind map. Lastly, by asking hypothetical questions students could make educated guesses and informed predictions with the help of the mind map. Asking hypothetical questions not only strengthened the various inferences made but also allowed students to explore the possibility of other claims so as to further support the claim of their choice and refute the other claims.

In one sense, the cognitive processes (see Fig. 1) that group F had undergone were similar to how authentic crime scene investigation was conducted in real life where questioning, hypothesis building, verifying or eliminating the hypotheses were integral parts of the investigative procedures. We recognized that this model was the product of a limited study in one context that was based on a case-study of a group that yielded the highest score for the quality of the argument they constructed. Nonetheless, it provided us with insights on how questions could be orchestrated to facilitate the construction of quality arguments. In addition, the model was also predictive, providing a framework for future research. The model suggested that cognitive processes employed this approach of using questions allowed for more valid inferences and encouraged students to consider multiple perspectives during problem solving and argumentation. This on its own could be viewed as a strategy in the construction of arguments.

Conclusion

We want to make three key assertions from the observations of this study that would help us better understand the conditions for using questions in quality argumentations. Firstly, from the study conducted by Chin and Osborne (2010a), groups that asked more questions specifically on key inquiry and basic information were better at generating quality arguments. The findings from this study concurred with that of Chin and Osborne (2010a). Secondly, our results provided an insight to hypothesis questions, another category of questions not discussed by Chin and Osborne in their study. Asking questions that are hypothetical in nature promoted the construction of argumentation from multiple perspectives and triggered fellow discussants to consider propositions to support a claim and refute others. Specifically, for the more successful groups, they

adopted strategies such as visible schema constructed from questioning, opting for the claim with the highest number of propositions and opting for claim that can be supported but not refuted.

Thirdly, this study revealed that although the types of questions mattered in the production of sound and strong argument, it was how these questions were organized into a visible schema that further augmented the development of quality arguments. The categories of questions discussed in this study had a mutually symbiotic relationship. Key inquiry questions could not be examined with much validity unless there was sufficient basic information. The volume of basic information depended on the number of questions seeking these basic information. On the other hand, hypothetical questions that often led to a spin-off of other perspectives depended on the pool of basic information. Though not revealed in the present study, it might be possible that students would not find answers to their hypothetical or key inquiry questions. They might resort to asking more basic information questions to help them construct premises and inferences for their key inquiry or hypothetical questions. In this way, one could see the interdependence of these three types of questions and the role they played in the development of quality argumentation.

Authors' contribution

MP was involved in the data collection, analysis and writing. ALT was involved in checking the analysis as well as the writing of the manuscript. MP and ALT would like to thank the students and the school for their involvement in the project. Both authors read and approved the final manuscript.

Funding

This project is individually funded by May and Aik-Ling.

Availability of data and materials

The data is not available for sharing due to the confidentiality clause included in the informed consent for the study.

Competing interest

We declare that we have read through the areas of competing interest, both financially and non-financially and declare that we do not have any competing interest. The data collected from this project has obtained the necessary clearance from the school and the students involved in the study.

Author's information

May Phua is a secondary school biology teacher who did her masters in education with Aik-Ling Tan. Aik-Ling Tan is a science education researcher with the National Institute of Education, Nanyang Technological University, Singapore.

Publisher's Note

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

Author details

¹Methodist Girls' School, 11, Blackmore Drive, Singapore 599986, Singapore. ²National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616, Singapore.

Received: 29 June 2017 Accepted: 29 January 2018

Published online: 15 March 2018

References

- Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*, 765–793.
- Chin, C. (2004). Students' questions: Fostering a culture of inquisitiveness in science classrooms. *School Science Review*, *86*, 107–112.
- Chin, C. (2006). Using self-questioning to promote pupils' process skills thinking. *School Science Review*, *87*, 113–122.
- Chin, C., Brown, D. E., & Bruce, B. C. (2002). Student generated questions: A meaningful aspect of learning science. *International Journal of Science Education*, *24*, 521–549.
- Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning science. *Studies in Science Education*, *44*, 1–39.
- Chin, C., & Osborne, J. (2010a). Students' questions and discursive interaction: Their impact on argumentation during collaborative group discussions in science. *Journal of Research in Science Teaching*, *47*, 883–908.
- Chin, C., & Osborne, J. (2010b). Supporting argumentation through students' questions: Case studies in science classrooms. *The Journal of the Learning Sciences*, *19*, 230–284.

- Chinn, P. W., & Hilgers, T. L. (2000). From corrector to collaborator: The range of instructor roles in writing-based natural and applied science classes. *Journal of Research in Science Teaching*, 37, 3–25.
- Curriculum Planning and Development Division [CPDD]. (2013). *Science syllabus – Lower secondary*. Singapore: Ministry of Education.
- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, 12, 91–142.
- Duschl, R. A. (2007). Quality argumentation and epistemic criteria. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspective from classroom-based research* (pp. 159–175). Dordrecht: Springer.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argumentation pattern for studying science discourse. *Science Education*, 88, 915–399.
- Harper, K. A., Etkina, E., & Lin, Y. (2003). Encouraging and analyzing student questions in a large physics course: Meaningful patterns for instructors. *Journal of Research in Science Teaching*, 40, 776–791.
- Hurd, P. D. (1998). Scientific literacy: New minds for a changing world. *Science Education*, 82, 407–416.
- Jimenez-Aleixandre, M. P., & Lopez, R. R. (2001). Designing a field code: Environmental values in primary school. *Environmental Education Research*, 7, 5.
- Jimenez-Aleixandre, M. P., & Pereiro-Munoz, C. (2005). Argument construction and change while working on a real environment problem. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 419–431). Dordrecht: Springer.
- Jimenez-Aleixandre, M. P., Rodrigues, A. G., & Duschl, R. A. (2000). 'Doing the lesson' or 'doing science': Argument in high school genetics. *Science Education*, 84, 757–792.
- Jonassen, D. H. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environment*. New York, NY: Routledge.
- Kelly, G. J., Chen, C., & Prothero, W. (2000). The epistemological framing of a discipline: Writing science in university oceanography. *Journal of Research in Science Teaching*, 37, 691–718.
- Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, 86, 314–342.
- Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education*, 83, 115–130.
- Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, 36, 1065–1084.
- Kolsto, S. D. (2006). Patterns in students' argumentation confronted with a risked-focuses socio-scientific issue. *International Journal of Science Education*, 28, 1689–1716.
- Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education*, 25, 1387–1408.
- Manson, L. (1996). An analysis of children's construction of new knowledge through their use of reasoning and arguing in classroom discussions. *International Journal of Qualitative Studies in Education*, 9, 411–433.
- McNeill, K. L. (2011). Elementary students' views of explanation, argumentation and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, 48, 793–823.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15, 153–191.
- Mercer, N. (2000). *Words and minds: How we use language to think together*. London, England: Routledge.
- Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead, UK: Open University Press.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41, 994–1020.
- Pfundt, H., & Duit, R. (1994). *Bibliography: Students' alternative frameworks and science education* (4th ed.). Kiel: Institut für die Pädagogik der Naturwissenschaften (IPN).
- Prain, V., & Hand, B. (1999). Students' perceptions of writing for learning in secondary school science. *Science Education*, 83, 151–162.
- Rivard, L. P., & Straw, S. B. (2000). The effect of talk and writing on learning science: An exploratory study. *Science Education*, 84, 566–593.
- Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92, 447–472.
- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23, 23–55.
- Sandoval, W. A., & Millwood, K. A. (2007). What can argumentation tell us about epistemology? In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspective from classroom-based research* (pp. 71–88). Dordrecht: Springer.
- Tan, A-L., Lee, P. P. F., & Cheah, Y. H. (2017). Educating science teachers in the twenty-first century: implications for pre-service teacher education. *Asia Pacific Journal of Education*, 37(4), 453–471.
- Vygotsky, L. S. (1962, 1986). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walton, D. (1996). *Argument structure: A pragmatic theory*. Toronto, CA: University of Toronto Press.
- Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, 79, 1–28.
- Zohar, A. (2007). Science teacher education and professional development in argumentation. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspective from classroom-based research* (pp. 245–268). Dordrecht: Springer.
- Zohar, A., & Nemet, F. (2001). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39, 35–62.