The Journal of Educational Research

ISSN: 0022-0671 (Print) 1940-0675 (Online) Journal homepage: https://www.tandfonline.com/loi/vjer20

Improving teacher questioning in science using ICAP theory

Joshua Morris & Michelene T. H. Chi

To cite this article: Joshua Morris & Michelene T. H. Chi (2020) Improving teacher questioning in science using ICAP theory, The Journal of Educational Research, 113:1, 1-12, DOI: 10.1080/00220671.2019.1709401

To link to this article: https://doi.org/10.1080/00220671.2019.1709401

Published online: 31 Jan 2020.

Submit your article to this journal

Article views: 111

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=vjer20

THE JOURNAL OF EDUCATIONAL RESEARCH
2020, VOL. 113, NO. 1, 1–12
https://doi.org/10.1080/00220671.2019.1709401

Routledge
Taylor & Francis Group

Check for updates

Improving teacher questioning in science using ICAP theory

Joshua Morris and Michelene T. H. Chi

Learning and Cognition Lab, Payne Hall 122, Arizona State University, Tempe, Arizona, USA

ABSTRACT

The purpose of this research was to investigate the efficacy of professional development in changing two middle school science teachers' questioning to include more questions that require deeper student responses. The professional development was based on ICAP theory which proposes a framework for identifying cognitive engagement based on what is required of students. ICAP hypothesizes that Interactive and Constructive questions, such as those requiring inferential thought and collaboration, lead to deeper thinking and therefore stronger learning gains than questions that are more Active, Passive. Teacher questioning before and after the PD showed a marked improvement in increasing the number of questions that required students to make inferences about the materials. Student gain scores from tests on the lesson's content suggests a positive relationship between *constructive* questions and student learning. This paper also details a coding scheme based on ICAP that can be used to provide feedback on teachers' questioning.

ARTICLE HISTORY
Received 1 May 2019
Revised 17 December 2019
Accepted 22 December 2019

KEYWORDS
ICAP; professional development; questioning; science teaching

Introduction

Changes to standards have increased the expectations for active learning during science instruction (National Research Council, 2012), such as allowing students the chance to explore their ideas through talk with each other and the teacher (Windschitl & Stroupe, 2017). One way science teachers can meet this expectation of scaffolding deeper understanding during classroom discussions is through improving the questions they give to their students (Lee & Kinzie, 2012; Monk & Osborne, 2000; Resnick, Michaels, & O'Connor, 2010). Unfortunately, it is well reported how infrequently teachers ask questions that require students to elaborate and move beyond the material provided in order to enact deep cognitive engagement with the science content (Banilower et al., 2013; Benedict-Chambers, Kademian, Davis, & Palincsar, 2017). Specifically, teachers rely mostly on *low-level* questions that require students to recall discrete bits of knowledge (Benedict-Chambers et al., 2017; Eshach, Dor-Ziderman, & Yefroimsky, 2014). Because of the mismatch between teacher practice and science standards, particularly related to crucial concepts such as inquiry (Davis, Petish, & Smithey, 2006) and the nature of science (Abd-El-Khalick & Lederman, 2000), there is considerable interest in designing professional development to increase the quantity and quality of constructivist practice through improving teacher questioning.

Designing effective professional development for questioning

There are a number of descriptive studies observing the effective ways that teachers engage their students with questions, such as encouraging students to make predictions (Lee & Kinzie, 2012) justify their reasoning (Chen, Hand, & Norton-Meier, 2017) clarify their thinking (Van Zee & Minstrell, 1997) and connect concepts to personal experience (Van Zee, Iwasyk, Kurose, Simpson, & Wild, 2001). Some of this work provides advice for how teachers can use particular talk moves for improving the variety and depth of their questioning (Chin, 2007; Michaels & O'Connor, 2015). There is, however, a surprising lack of research on effective methods for training teachers how to use questions in ways that encourage students' cognitive engagement. In an older review of training studies, Winne (1979) concludes that many studies exploring the connection between training teachers to ask questions and student achievement, described the training, and measured student learning, but failed to measure whether the training had actually improved teachers' questioning practice in the classroom. This lack of data makes it difficult to determine the efficacy of teacher training on improving teacher questioning in the classroom. More recently there have been some promising studies that have monitored the change in teacher questioning behavior following extensive professional development.

Oliveira (2010) trained teachers on how certain questions can indicate to students that their contributions are central to the inquiry process. For example, questions about what a student observed signals to the student that their ideas and curiosities are important. In their professional development on student-centered questioning, the authors focused on questions that elicited student personal-experiences and questions that helped students, "derive more refined meaning from their own individual experiences" (p. 424). The PD involved teachers reading research illustrating the value of

CONTACT Joshua Morris morri50@asu.edu Learning and Cognition Lab, Payne Hall 122, Arizona State University, Tempe, AZ, USA.
© 2020 Taylor & Francis Group, LLC

inquiry-based teacher questioning. The researchers also gave examples of types of questions that aligned the teacher with a student-centered approach. For example, the researchers focused on *referential* questions through which teachers could learn about students' personal experiences or opinions about the subject (e.g., What do you notice about this rock?) as compared to a *recitation* question (e.g., "Who remembers what we call this type of rock?") or a close-ended question (e.g., "Who knows what type of rock this is?"). The goal of the student- centered approach to questioning was that it indicated the teachers' open attitude toward student ideas and observations. Three of the participating teachers agreed to be videotaped and these teachers demonstrated a positive change toward more student-centered questioning and the researchers attributed this change to the PD's focus on drawing connections between the proposed changes in practice to teacher's experiential knowledge (Borko, 2004; Desimone, 2009).

In their study of questioning in pre-k classrooms, Lee, Kinzie and Whittaker (2012) sought to increase teachers' open-ended questioning in order to facilitate the students' linguistic development. *Open-ended* questions were defined as a question with multiple acceptable answers, as compared to a *closed-ended* question that has only one acceptable answer (Hargreaves, 1984). The reasoning put forth by the authors was that open-ended questions require more linguistically complex answers and this should lead to increased and more diverse student talk. The professional development program included general teaching tips about instructional strategies and student thinking, as well as a video review of participating teachers' self-recorded instruction where researchers highlighted instructional sequences that could foster open-ended questions. Using an experimental design, the authors showed that their program of professional development increased teachers' open-ended questions and that this increase was related to an increase in the syntactic complexity of student answers (e.g., number of clauses).

In their Questioning the Author (QtA) approach, Beck, McKeown and colleagues (Beck, McKeown, Hamilton, & Kucan, 1997; Beck, McKeown, Sandora, Kucan, & Worthy, 1996) present teacher *queries* as a type of question that is not focused on recitation of student knowledge but instead is designed to facilitate the elicitation and discussion of student ideas. In the QtA approach a good query is one that results in students building on their peers' ideas, or drawing inferences based on the information provided in the text. The central component to the QtA approach is student engagement with a text of some kind. Most of the authors' recommended queries rely on the students comparing and contrasting their understanding with the `intention of the author, and the text as written.

Overall, PD based on QtA decreased the number of teacher questions focused on simple recall and increased the number of questions that where students drew inferences or reflected on other students' comments. Student learning was measured as improvement in their comprehension defined as the students' ability to provide an answer to the question "What was the author's message?" Pre- and post-intervention assessments showed that students' participation in QtA improved their ability to answer this question. It is important to note that most of the authors' recommended questions rely on the students comparing and contrasting their understanding with the intention of the author, and the text as written.

All of these studies were attempting to decrease teacher questions that required students to recall or remember discrete facts from a text or memory and increase the number of questions that the authors deemed to be of a higher caliber. In Oliveira (2010) the authenticity of teacher questions (e.g., questions about opinion, observations) was paramount as a means of indexing the importance of the students' engagement with knowledge building. While their results showing increased use of questions that elicit student experience is valuable to inquiry process in science, there is still a need to demonstrate how best to train teachers to ask questions that elicit student responses that require deeper engagement with the material not just responses focused on student experiences.

Both Oliveira (2010) and Lee et al. (2012) also draw much needed attention to student discourse around the classroom content, but there is still a need to explore connections between teacher instruction and individual measures of learning.

Beck et al. (1996) set out to increase questions that focused students on drawing inferences about authorial intent in order to better comprehend. Questioning strategies centered on a text are limited as teachers may be discussing science content in the absence of a focal text, and they may need a different framework through which to analyze and plan their questioning.

In all of these approaches, there are various classifications for what warrants a good question. In this study we will be using ICAP theory for helping teachers organize various types of questioning methods into one framework. The goal is to document changes in teachers' questioning patterns following a theoretically-grounded but relatively brief professional development workshop featuring a parsimonious framework for fostering cognitive engagement.

The theoretical basis for this PD will be ICAP (Chi, 2009; Chi & Wylie, 2014). The ICAP theory provides an organizing framework for tracking cognitive engagement by monitoring students' overt behaviors and students' products during the learning activities. The theory suggests that levels of engagement can be reasonably inferred based on the observable actions of the students during learning activities as well as their products (which can be notes and/or diagrams students wrote, or their questions and comments, etc.) In this way, ICAP can serve as a tool for analysis of teachers' questions as well as helping the teachers assess cognitive engagement on the fly.

ICAP proposes there are four engagement modes that can be used to categorize student cognitive engagement: Interactive, Constructive, Active, and Passive. Teachers can design activities that are more or less likely to elicit a level of engagement. When a teacher lectures and does not ask students to do anything with the material being presented (e.g., note taking), the students are most likely *passively* engaged. The essential feature of passive engagement is the absence of any overt behaviors to act on the information and/or absence of any products resulting from the overt behaviors. If the teacher makes requests of the student to use the material in some way such as copying notes from the board, then they are likely *actively* engaged because the notes contain information similar to the teacher's notes. If students are asked to go beyond the material in some way requiring them to make inferences then the students are being *constructive*. The key distinction between being actively and constructively engaged is that constructive engagement requires going beyond the information as provided. Examples of constructive teacher requests would be asking the students to create concept maps where they are required to add connecting words and justify the organization, or rephrasing substantive portions of presented material in their own words. An *interactive* learning activity would be similar to a constructive activity but one where students have to work collaboratively to complete. An interactive task is one that requires using each other's contributions to co-infer new ideas, generating questions and providing each other with answers, or engaging argumentatively over a controversy. Importantly, cooperation is necessary but not sufficient for interaction; the students must be building on one another's contributions.

ICAP can also be used as a tool for generating questions in activities and classroom discussions. In essence, the practitioner that uses ICAP to monitor their question is looking for 3 qualitative shifts in their students' activity: From listening to engaging with the material, from engaging with material to moving beyond the material, from moving beyond the material independently to co-inferencing with peers. In terms of teachers generating questions during classroom instruction, the most important shift is the second; reducing questions that likely result in students engaging in shallow processes such as identifying or recalling and increasing questions that result in deeper more inferential processes. In ICAP this can be conceptualized as the shift from *active* question to *constructive* questions.

For teachers generating questions on the fly, the concept of a *Constructive* question is useful in its parsimony. According to ICAP the essential characteristic of a constructive question is whether it requires students to go beyond the presented material in order to answer and this can be easily translated into a guiding heuristic for teacher.

We contend that, the concept of ICAP and the defining features of *constructive* questions specifically, are easier than other question typologies for teachers to utilize. The main reason is that ICAP provides an easily remembered conceptual anchors to follow when they want questions that elicit deeper thinking. As long as teachers can focus on the guiding question, "Does this question ask the student to go beyond the material that has been provided so far?" they will have a greater chance of students engaging in elaborative and generative thinking.

Previous work on professional development using ICAP to increase student cognitive engagement shows that teacher knowledge about constructivist practice improved, but the activities and questions teachers implemented were not effectively designed to elicit higher- levels of student thought (Chi et al., 2018). In this previous work, teachers were allowed to design their own activities and lesson plans for a content area of their choosing whereas in the current study the participating teachers will be teaching the same science lesson. Another difference between this study and the previous work is that the PD for the prior project did not focus on question-asking per se, which is a focus in the current PD.

The current study improves on the PD used in previous studies in a few ways. The changes most relevant to improving teacher questioning were the following: (1) Addressing teacher misconceptions about constructivist practice, (2) Drawing connections between ICAP theory and teachers' current practice, (3) Providing examples of question types teacher use to foster deep thinking, (4) Providing videos of teacher questioning behaviors that are more and less constructive and (5) Reviewing lessons to allow teacher to brainstorm where constructive/interactive engagement would be most fitting. We provide more detail about these changes in the methods section, *Professional Development.*

Research questions and study design

Prior research on ICAP has shown that teachers value and understand the ICAP framework, but still struggle and guiding discussions that demand students go beyond previously learned content (Chi et al., 2018). The current study's PD has been elaborated with this finding in mind, mainly, by providing the lessons and activities we were controlling for the variability in teacher lesson design and focusing on whether we could increase constructive and interactive questioning.

The research questions for the current study are,

1. Does professional development centered on teachers' questioning based on the ICAP framework increase the amount of constructive and interactive questioning used in teachers' instruction?
2. Can the ICAP framework be used to provide a straightforward coding scheme for questioning used during instruction?

To answer these research questions, we recruited two 7th grade science teachers to participate in professional development about how to use the ICAP framework for analyzing and planning the questions used in their instruction. The teachers' instruction was recorded before and after their participation in the ICAP PD. There were two lessons recorded following the PD. Following the instruction of the two lessons, we compared the teachers' instruction after receiving PD to their own instruction prior to PD.

Materials and methods

Participants

The data comes from 2 different 7th grade teachers who each taught two lessons. The two teachers were recruited for this study to test the current iteration of a professional development module and to pilot two lessons designed to be used in the 8th grade curriculum in a participating district. Both teachers came from the same school in a large urban district in the southwest. The science department head for that school facilitated the recruitment and participation of the teachers who were paid for their participation.

The participating students were 7th graders recruited from the participating teachers' science classes. The teachers and science department head were asked to recruit a sample of students that contained a typical class' mixture of giftedness, gender, race, talkativeness, interest in science, and students who had an Individualized Education Program. The lessons were taught on a Saturday, and so students were incentivized with extra credit for their class, and a pizza. lunch provided by the researchers. In total, 21 students agreed to participate and were split 11 and 10 between the two teachers, with attention paid to the distribution of student characteristics listed above. Five of the participating students had an individualized education program and three were placed in one of the classes, and two in another. Four of the students had been identified by the district as gifted and they were split evenly between the classes. Other informal classifications made by the teachers (e.g., "quiet", "can get off task") were noted and used by the researcher to construct congruent groupings. The lesson and observation took place over the weekend because the lessons were not part of the 7th graders' normal curriculum. All students provided consent to be a part of data collection.

Participating students completed two pretests (one for each lesson's content) prior to instruction. They then received one lesson (~1 hour) in the morning, the first posttest, a lunch break, the second lesson (~1 hour) and then the second posttest. There was no time limit and most students finished each test in around 15 minutes. Both posttests were given within thirty minutes of finishing the lesson.

Lessons

Each teacher taught two lessons. One was on genetics and the other was on adaptations. The teachers were instructed to teach one lesson in a constructive manner and the other lesson more interactively. Lessons were counterbalanced so that each lesson was taught in each of the two engagement modes (see Table 1).

To answer the first research question, the pre- and post-PD recordings of teaching was transcribed and coded for question types. Any differences in question types served as a metric for changes in the teachers' instruction. With the counterbalanced design, we were also able to compare the questions they asked in each of the engagement mode conditions (e.g., Interactive or Constructive) which provided an indication of the teachers' ability to enact these two higher-levels of ICAP. The biggest difference we expected to find is that teachers would use more active and passive questions in their baseline lessons and more constructive and interactive questions in the lessons following the PD. We also expected that during each teachers' interactive lesson they would provide students more opportunities to interact with their peers.

Table 1. Content of lesson taught by teacher in each engagement mode.

	Interactive lesson	Constructive lesson
Teacher 1	Genetics	Adaptations
Teacher 2	Adaptations	Genetics

To answer the second research question, this study provided an opportunity to refine a coding scheme (piloted in a previous study) to code teacher questions using the ICAP framework. One key component in the ICAP framework is that a question is only considered enact the highest levels of processing (e.g., constructive and interactive) if the answer requires the student(s) to beyond the material as presented. Other coding schemes categorize questions based on the cognitive process necessary to produce an answer, but without a consideration of the context in which the question was answered, assumptions about cognitive engagement could be specious (Winne, 1979; Lee & Kinzie, 2012). We contend that simply attending to relevant contextual elements (e.g., what is printed on lecture slides) will add to our coding scheme's efficacy.

We asked the teachers to select a lesson before the professional development in order to provide a baseline for their teaching practice. They were asked to pick a lesson representative of their teaching. Specially, a lesson that they felt, (1) would likely serve as a representative sample of classroom discourse, and (2) included some student activities. Examples such as, *amount of teacher talk, opportunity for peer collaboration*, were provided to clarify what was meant by *classroom discourse*. The intention was that, once transcribed, this lesson would serve as a baseline for the types of questions these teachers tended to ask during a lesson that involved some student activities. Both teachers selected a Geology lesson wherein students were classifying rock samples.

The two lessons taught after the PD were selected because they had been created for use in a study in the same district and using them in this study provided pilot data for the larger study. The two lessons were modifications of existing lessons in the district's 8th grade fall curriculum. As mentioned above (Point 5, in section on improvements made to PD), an integral part of our professional development was having teachers apply concepts from ICAP theory to the lessons they would teach. Since the modules were tied closely to these 8th grade lessons we wanted the participating teachers to use these lessons. This also gave us the chance to observe teachers applying the PD to a lesson they hadn't taught before. The advantage being that the PD could have maximum effect if teachers did not have interference from any pedagogical content knowledge that might influence how they applied ICAP to the lessons. Such data could serve as a baseline for the PD's effectiveness in future studies.

One lesson was on genetics and centered on a "create-a-kid" activity where students were tracking alleles passed from parents to child in order to identify the child's phenotypes. The second lesson was on adaptations and centered on an activity where students were tracking changes in a pocket mouse population due to changes in the environment. The participating teachers did not design these lessons nor had they taught these lessons before. They were, however, familiar with the content.

Professional development

The professional development module was created with an emphasis on the five modifications based on previous modules and teacher misunderstandings detailed in previous work (Chi et al., 2018) and introduced above. First, the PD elicits teacher beliefs about constructivist practice and provides counterarguments to commonly held naïve conceptions about learning (Tobin & McRobbie, 1996). Data from studies measuring both beliefs and instruction suggests that instructional practices such as asking mostly low-level questions are tethered to a deeper belief about how knowledge is constructed, and so attempting to change the action without addressing the belief would be ineffective. (Brickhouse, 1990; Hashweh, 1996; Tobin & McRobbie, 1996). This was an addition to this PD because of our interest in studying teacher questions during instruction, where teachers are most likely to rely on fact level questions (Masalmeh, 1998).

Second, the professional learning literature recommends that there needs to be coherence between the content of the PD and the teachers' materials and curriculum (Desimone, 2009).

This was generally achieved by highlighting how ICAP was related to overall trends in education such as the importance of inquiry in science learning. More specifically, certain district initiatives were linked to ICAP, such as their adoption of AVID's WICOR framework to foster inquiry in the sciences (www.AVID.org). Specifically, during the PD teachers were shown how the I and C of ICAP aligned with the I(inquiry) and C(Collaboration) or WICOR.

Third, as shown in the reviewed studies it is advantageous to provide concrete examples of the types of questions that teachers can use to elicit constructive responses and facilitated interaction. We focused on the types of questions teachers could use during whole-class discussion. Similarly, teachers watched video examples of other teachers and practice determining which questions were more and less constructive.

Fourth, we conducted the PD with teachers in one district which facilitated the opportunity for authentic collaboration and connection between the theory and the curriculum (Garet et al., 2004).

Finally, in the last stage of the PD, instead of having each teacher work independently on creating their own lessons, teachers used researcher created lesson plans and, along with the moderator, brainstormed which areas of the lesson would lend themselves to more *constructive* and *interactive* processes. Primarily, where to insert questions that elicited constructive and interactive processes. This feature of the PD was in direct response to prior work showing that teachers designing their own lessons after learning about ICAP had trouble designing activities that led to *constructive* and *interactive* processes. By providing the lessons, teachers would be able to focus on the questioning and instructional moves that led to constructive processes. Importantly, teachers were not provided with a script or specific questions to ask students.

The PD was given over the course of two days: One Saturday (7 hours) and a Wednesday after school (4 hours). This was done to accommodate the two teachers' schedules. Teachers were guided through the module with a mixture of lecturing with Powerpoint slides and interactive activities.

The content of the professional development (PD) module was divided into four sections: 1. Introduction to ICAP, 2. How to Scaffold Constructive Processes, 3. How to Scaffold Interaction 4. Teaching the two target lessons. For section 1, students were introduced to how ICAP improved upon vague educational imperatives for "active learning." The acronym was defined and examples were given of processes for each engagement mode. Less emphasis was put on the underlying knowledge processes than in previous modules.

For section 2, teachers were introduced to best practices for how to elicit constructive responses in the classroom. Teacher were introduced to discourse moves that could facilitate classroom discussions (e.g., wait time, framing the activity, provision of materials) but the section was primarily focused on teacher questioning. Examples of questions that were more and less *constructive* were provided with the emphasis that that question stems were helpful but not sufficient indicators of engagement and, instead, the teacher must consider the information available and previously presented. In this lesson teachers were exposed to the research concept of "dosage" with the explanation that their goal was to increase the number of constructive and interactive questions and opportunities as opposed to drastically altering the lesson plan to be entirely problem- or inquiry-based. This was intended to counter the common misconception that inquiry learning is orthogonal to any recitation or lecturing. Instead, we emphasized that the goal is to operate within the lesson plans they already have in place but to increase the probability of students' higher cognitive engagement by offering constructive questions and providing opportunities for the students to move beyond the material as presented.

For section 3, teachers were introduced to best practices for how to scaffold interaction. Refutational texts were used to address common misconceptions about collaborative processes. Teachers were presented with common maladaptive group dynamics and ideas for how to deal with such situations. Transcripts of student talk that were more and less interactive were presented and discussed. Videos of exemplary and poor scaffolding of student interaction and teacher questioning were shown and discussed.

For section 4, the teachers reviewed the provided lessons that they were going to teach. Importantly, the lessons were not scripted and teachers were left to generate their own questions and guide the discussions.

While working through the first day of the module there were multiple interactive components including (a) Using ICAP to self-assess the teacher's usual instructional practice for engagement modes using ICAP framework (b) Sorting classroom tasks by ICAP engagement mode (c) identifying engagement mode using work products (d) re-designing a science lesson (not the lessons they were to teach) to increase constructive and interactive processes (e) reading and reacting to refutational texts identifying common misconceptions about engagement mode in the classroom (f) identifying engagement mode through monitoring student talk in whole class discussions and in concurrent collaborative groups (g) watching and analyzing videos of classroom instruction for evidence of interactive, constructive, active, and passive engagement.

In regards to focusing on teacher questioning, as part of the lesson review (d, above) teachers identified constructive questions that they could interject into the lessons to help students think about key concepts or teacher-identified sticking points. This process of planning questions served a pedagogical function for the research team as it provided an opportunity to give teachers feedback on their questioning. In watching and analyzing of videos (g, above) teachers were presented with transcripts of the questions that were asked during videos of different science lessons. Teachers used the context of the lesson and wording of the questions to determine whether they served a passive, active, constructive, or interactive function. The three videos were chosen to demonstrate teachers' use of mostly active, mostly constructive, and mostly interactive questioning.

On the second day of the PD, the two lessons were presented to the teachers with an emphasis on the pacing of the lesson and materials. It was made clear to the teachers that the goal of the study was to see how they used ICAP to inform their instruction and questioning. To this end, scripts were not provided, and questions were not provided beyond those included in the Powerpoint materials. The provided slides introduced the essential understanding for the lesson, introduction to the topic and an outline for the activity and any corresponding instructions. For both lesson there were 8 questions on the provided slides that could be asked during classroom discussion (constructive condition) or discussed interactively (interactive condition). How the questions were used was left up to the teacher. By providing some guiding questions and the outline of the activity the provided lesson was meant to serve as a starting place for the participating teachers with the expectation that they would flesh out the lesson using what they learned in the professional development. For example, similar to the lesson re-design activity in the first day, teachers brainstormed about areas where constructive questioning and interactive collaboration would be fitting. Some overarching essential questions were identified by the teachers. Because the lessons were to be taught in predominantly one mode or another (e.g., *Constructive* or *interactive*, see "Study design" above) the teachers also discussed specific activities that could be completed either using collaborative groups for the interactive condition, and with teacher-led discussion in the constructive condition. In order to control for content coverage, teachers were instructed to work within the lessons as presented, or to synchronize the changes across the lessons so that every student had access to the same activities. The clarify, teachers in both conditions worked from the Powerpoint slides and so part of their lesson was delivered via lecture. Their assignment to condition determined how they could modify their lessons. Where the primary changes involved students in the *constructive* condition being asked inferential questions or being given an activity where they had to generate some new thinking, and students in the *interactive* condition being allowed to answer the same questions or complete the same activities while working in pairs. The teachers had to modify the lessons jointly so that a change in one lesson was mirrored with an appropriate change in the other condition.

The assessments used to measure student learning were created by the research team. To design the assessments, we gathered assessment items from five state standard assessments as well as four available concept inventory items related to the NGSS Disciplinary Core Ideas for Life Sciences. Using difficulty ratings provided with the state assessments the assessment was crafted such that it contained 30% easy questions, 40% medium-difficulty questions, and 30% hard questions.

Following the teaching of the lessons, the teachers were interviewed about the usefulness of the ICAP framework in guiding their instruction and about the professional development in general. These interviews were transcribed.

Data analysis

Teacher questions were first transcribed from the 6 lessons (1 baseline, 2 treatment for each teacher) by two undergraduates. A question was defined as a statement with interrogative force and/or containing common grammatical elements of a question (e.g., begins question with what, when, etc.,). During the transcription process questions were tagged as follow-up questions if the teacher's question was in response to or extending from a students' statement.

After the transcription was completed, to identify whether teacher questions had been reliably identified, the videos were reviewed by whichever undergraduate had not done the initial transcription. Of the questions identified across the 4 videos, there was discrepancy on 8% of utterances identified as questions. The majority of these discrepant utterances involved teachers asking rhetorical question. The protocol was modified to make it clearer that such questions were to be transcribed. Overall the percent agreement for question identification was initially 92% and 97% with the modified protocol. For follow-up questions, there was disagreement on 6% of the questions identified as follow-up. These disagreements were resolved through discussion between the two coders.

The creation of the coding protocol largely followed the eight steps laid out for analyzing verbal data in Chi (1997). The ICAP framework provided the formalism upon which the four primary codes were based and given the framework's thorough development in Chi and Wylie (2014), we were able to create clear parameters for the coding scheme based on the content of the question and the information provided in the lesson. Coding of questions was carried out in two phases. All coding was done by the first author and an undergraduate lab member. In the first phase, the undergraduate was taught operational definitions of each ICAP engagement mode (as outlined in section, "Coding scheme"). The undergraduate was then trained on transcribed questions from vides of instruction taken from the internet and previous studies. A trial phase involved the undergraduate and first author coding the same questions separately and then convening to discuss any discrepancies or questions. Two trials were needed to obtain satisfactory reliability.

In the second phase, the undergraduate independently coded all transcribed questions from all lessons. Although the questions were previously transcribed coding was done while watching the video of the lesson in order to note any important contextual factors that had a bearing on coding. As we will discuss below, the contextual factor most important to our coding was whether the teacher had provided the students with the information necessary to answer the question. The first author coded 40% of the total number of questions (a sample from every lesson and every teacher) to determine reliability. There was 92% agreement between the two raters.

ICAP as a tool for a coding scheme

Every question was coded as belonging to one of the four modes outlined in the ICAP framework: *Interactive, Constructive, Active* or *Passive*. One unique aspect of our coding scheme is that it is designed to account for information that was available to the students in the learning context. We constrained context to refer to information explicitly provided during that days' lesson, either orally, by the teacher or students, or visually via students' individual learning materials or Powerpoint. No assumptions were made about student background knowledge as they would be largely untenable.

Questions pertaining to classroom protocol or procedural concerns were neither coded nor used in calculating percentages.

Passive questions

Questions coded as "passive" were those that resulted in no expectation for the student to verbalize an answer. The most common type of question that was coded as passive was one where the teacher provided the answer to their own question, in a follow-up question. An example from our sample would be,

> Teacher: *So … where's the gene? It's on the chromosome, right?*

Obviously, the timing between the first and second question might alter the coding, but there were few instances where this was an issue. In most cases the teachers asked the question and then answered it, within a few seconds. The second most common question that was coded as passive were those where the question's referent was vague and student responses were nods or shakes of the head. An example is,

> Teacher: *Does that make sense?*

Active questions

Questions were coded as "active" if the students were asked to recall information from previously taught materials, or provide discrete facts from background knowledge. These questions were considered active because neither encouraged the students to draw inferences about the lesson materials and both focused on recall of discrete bits of information. We determined that something was previously taught if the answer was provided explicitly prior to the question, or if the teacher pointed out that it was something they had covered before.

Explicit reference

> Teacher: *Who remembers what we talked about yesterday with predator and prey, about the relationship between predator and prey?*

Discrete knowledge, no inference

> Teacher: *What color is lava?*

Some questions that were ostensibly inferential in nature, were coded as active if the answer was provided by the teacher, peer or within any of the lesson materials during that lesson. An example is illustrative.

In a lesson on natural selection, two teachers working from a shared set of powerpoints (i.e., identical in content) both asked one of the lesson's essential questions, "When can a mutation be beneficial?" Teacher 1, asked the question to the class after presenting two slides. One slide described gene mutation and included the sentence, "Some mutations can be beneficial depending on the environment." The second slide gives the overview of the issue,

> In New Mexico, there is a population of mostly light-colored pocket mice that thrive in their light-colored environment. Random mutations have caused a few of the mice to be dark-colored. There was a volcanic eruption resulting in parts of their environment becoming dark-colored. In this dark colored environment, the light-colored mice no longer have protective coloration.

Teacher 2 asked the same essential question prior to showing any slides, encouraging the students to make predictions based on their background knowledge.

These two questions, identical in form, are very different given the context in which they were asked. For students in teacher 1's class, the essential components of the answer were directly covered by the slides and her overview of the slides. The most likely answer provided by students in this situation would be a retelling of the facts that were provided. For students in teacher 2's class, the question requires a lot more prediction and creativity based on whatever knowledge they already have about how environments interact with populations. These answers are far more likely to include integration of background knowledge or creative predictions, both processes that involve going beyond the information provided. Teacher 1's question is active while Teacher 2's question is constructive.

Constructive questions

Questions coded as constructive were those that required students to go beyond presented material in order to answer. As outlined above, going beyond the material was coded using two necessary criteria. First, most requisite parts of the answer must not have been previously presented to the student and second, the answer to the question required an inference on the part of the student. We did not distinguish between levels of inference or the complexity of reasoning necessary for the inference. Inferences are required for all the questions in the following sample, regardless of whether the form of the question is asking for a prediction, an application, an explanation, a justification, and so forth:

What will happen to the mice that have the mutation described in the text?

Which parent is more likely to have a child with dimples?

Why do you think that's the case?

With all of these questions, the important point is that potential answers to the question have not been provided in the lesson prior to the question being asked. In other words, some of the question formats do resemble other coding schemes evaluating a questions requisite depth of engagement (such as Bloom's oft-cited taxonomy, Anderson & Krahtwohl, 2001), but we are not as concerned about the format or its resemblance to typically high-level questions. The thing that matters most in determining the ICAP mode of a question is whether it asks students to generate new ideas.

In our preliminary coding we found that there was some disagreement in whether an answer had been provided in the learning context prior to the question. For example, consider the example given above about Teacher 1 asking the question, "When can a mutation be beneficial?" A reasonable assumption on our part is that the most likely answer that a student would come up with that would satisfy teachers in this situation could be directly pulled from the information provided and so would be considered an *active* question. In this way, we erred toward a conservative coding of constructive question.

Interactive time

There are not many innately interactive questions commonly used in classroom discourse and these types of questions were exceedingly rare in our sample. Questions such as, "Does anyone disagree with, Mary?" are questions that cue certain dialogic processes that are infrequently used by teachers. Cataloging such questions, that explicitly cued interaction between students, could underrepresent how teachers used questions to encourage interactivity. Instead of looking for only interactive questions, we identified times when students were given constructive questions and told to work together. Interaction typically takes longer than students answering questions in a whole class discussion, and so we measured *interactive time*, the amount of time the students were given to work together on the constructive question, on top of noting how often such opportunities were given. Our reasoning was that the time provided for such interaction would be a better metric of interactivity than just the number of times students were encouraged to interact.

Results

Table 2 presents the percentage of total teacher questions in each lesson for the baseline and the two lessons. Each question was coded as either passive, active, or constructive questions. The baseline lesson is the lesson the teacher taught prior to the PD. The constructive lesson was the lesson where the teacher's goal was to elicit more constructive processes and the interactive lessons was the lesson where the teacher's goal was to elicit more interactive processes. Table 3 presents the breakdown of follow-up questions, there were no follow-up questions rated as passive or interactive so only totals and percentages for Active and Constructive questions are included.

The most striking result is that after the PD both teachers decreased the number of passive and active questions they used and increased their use of constructive questions and opportunities for interaction. This was the case for both initiating questions and follow-up questions. In line with the goals of the PD, the teachers included more interactive time in their interactive lessons. Interestingly, for both teachers the lesson they taught interactively included more total questions, and more constructive questions. Note that the content was different for each teachers' interactive lesson so this is not just a byproduct of lesson materials.

The two teachers had distinct ways of using questions in their baseline lessons. Teacher 1's instruction included a large percentage (55%) of recall questions which was primarily used to assess student understanding of the material after it was presented. These questions generally followed a brief lecture to introduce ideas, and the questions served as recitation of facts presented within the lecture. She would also use pseudo-questions to reiterate ideas already presented. For example, she might review a geologic concept by using the pseudo-question, "So, a sedimentary rock is formed from sediment, sand and smaller rocks, right?" These types of questions were coded as a *passive* question because the student neither needed to generate new ideas to answer this question nor did they need to recall any

Table 2. Percentage of total questions asked by the two 7th grade teachers for each ICAP mode.

		Total	Per minute	Passive	Active	Constructive	Interactive
Teacher 1	Lesson			%	%	%	Time
	Baseline	38	0.73	42	55	3	0
	Constructive	43	0.51	20	50	30	0
	Interactive	100	1.20	13	17	70	10:20
Teacher 2							
	Baseline	88	1.96	4	87	9	0:00
	Constructive	215	2.65	3	56	41	0:00
	Interactive	220	2.79	4	50	46	11:38

Table 3. Follow-up questions broken up by teacher and lesson.

	Baseline lesson		Interactive lesson		Constructive lesson	
Type of	Teacher 1	Teacher 2	Teacher 1	Teacher 2	Teacher 1	Teacher 2
Follow-up	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Active	7 (45)	14 (45)	22 (45)	33 (45)	7 (32)	49 (62)
Constructive	2 (55)	4 (55)	27 (55)	41 (55)	15 (58)	30 (38)
Total	9	18	49	74	22	79

Note. Follow up questions not included in Table 2.

information because the answer was provided in the question itself. Teacher 1's baseline lesson also provides an example of how context can determine the depth of a question. In this lesson the students were identifying rocks.

> Teacher: *Let's talk about it, what do you notice when you look at that rock.*
>
> Student: *Darker than the other… and shiny*
>
> Teacher: *If it's dark and glassy, what kind of rock do you think it is*?

In this exchange the teacher asks two questions. We coded the first question as active because the teacher is asking the students to report what they are seeing. The second question is potentially constructive because she is asking the students to engage in a inductive process of reasoning. In this context, however, the teacher is gesturing to an identification chart which lists the exact words *dark* and *glassy* as characteristics of igneous rocks. The question was coded active because the students weren't making an inference, instead they were engaged in more of a matching task.

The majority of teacher 2's questions (87%) were assessments of student knowledge and they were more integrated throughout the lecture compared to teacher 1. These *active* questions primarily required the recall of discrete facts that had been previously covered in the lecture or were provided on student materials (e.g., worksheets). The questions were asked in a hurried fashion and the teacher moved on after receiving the answer; seeking little elaboration Teacher 2 did not use any pseudo-questions as review.

Teacher one after PD

In the constructive lesson, Teacher 1 continued to use a high percentage of recall questions but the frequency of passive questioning dropped by half. The placement of questions did not change as the teacher still would lecture uninterrupted and use questions as assessment tools following the lecture. The change seen in both the lessons following the PD was the increase in frequency of constructive questions used in these mini-reviews following the lecturing. Instead of relying on more of the passive, almost rhetorical, questions (e.g., "But we learned that they both [rock types] time to form, right?") the teacher included more constructive questions to get students moving beyond the material. The excerpt below indicates this pattern of presentation of information followed by a constructive question.

> Some animals have a mutation that causes them to have no pigment and so they appear all white. This condition is called albinism… or we usually say the animal is an albino. You might have heard that word before. So here we have different pictures of animals that are albinos and you can see they're all white… this tarantula looks very strange. So what… let me ask you guys to get you thinking, *what might be a positive benefit from albinism, or being an albino, for some of these animals*?

In the interactive lesson, Teacher 1 doubled their total number of questions and the majority (70%) of these questions were constructive. The interactive time was the sum of time from 5 dyadic discussions about different constructive questions asked throughout the lesson. A good example of a constructive question used to spur interaction comes a few minutes after the excerpt above. "Now I want you to talk with your shoulder partner and discuss, what is an animal that isn't up here [indicating the Powerpoint] that would benefit greatly from the genetic mutation of albino… of being an albino?" This teacher used a timer to indicate how long the students could discuss. Their *interactive* discussions averaged two minutes each.

Teacher 1 also included many more follow-up questions with about half of them being constructive in nature. Most of these constructive questions involved teachers asking about the students' reasoning.

Teacher two after PD

In both of the lessons post-PD, Teacher 2 asked a higher volume of questions.

Noticeably, the frequency of the active questions stayed the same but her constructive questions increased. The inclusion of questions into the lessons was similar to her baseline lesson in that they were peppered throughout her lecture, as opposed to Teacher 1 who had blocks of lecture time and then question time. Teacher 2 used active questions to assess bits of knowledge and review concepts throughout the lecture as in the baseline lesson. The constructive questions were interspersed among the active questions. Similarly, in the interactive lesson, the constructive questions that the teacher asked dyads to discuss were included throughout the lesson. The interactive time was the sum of time from 7 dyadic discussions about constructive questions posed by the teacher. Some of these interactive moments were created after the students didn't answer a constructive question quickly enough. The teacher would have students do "think-pair- share" instead of providing additional wait time. In this way, it seemed that the teacher was using interaction as a way to stimulate thought on tough questions. These *interactive* discussions lasted around a minute and a half each.

Student learning gains

Table 4 shows the simple gain scores (e.g., post-test minus pretest) students in each of 7th grade classes. The bottom row (highlighted in dark grey) show the averages across the two classes. Both classes demonstrated learning across the two lessons. The small sample size suggests a cautious interpretation, but there is also some evidence that students in the interactive condition may have had a slight learning advantage over students in the constructive condition.

Teacher interviews

In the interviews with the two teachers two common themes emerged. First, both teachers preferred lessons taught in an interactive mode. Reasons they cited were that students seemed, "more engaged and willing to talk." and that, "everyone was caught up in the learning and answering."

Second, both teachers remarked upon the concept of dosage that had been highlighted in the PD. A discussion on teachers' questioning is illuminating,

> Interviewer: *Did you notice anything about your own questioning during the lesson?*

> Teacher 1: *The idea of that the questions were like doses really stuck with me. That I didn't have to make everything something… you know, deeper or conceptual but I just needed to keep asking more open questions… whenever it came up, when it made sense… To make a whole lesson collaborative and interactive is difficult but not as much… it's easier if it's only pockets of questions, you know, here and there.*

> Teacher 2 (answering the same question): *I tried to pay attention more to the students answers to see if they were just giving me back what they thought I wanted to hear. Like, that idea of 'going beyond' is hard… it's not hard… it's just easier when you listen to the answer. Then you can figure out if they're just guessing or regurgitating something, and you can help them think deeper with a follow up. Like we said, the doses of good questions… that was easier after the student said something and I could react. It was easier to change those.*

In both of the answers above, the teachers referred to the concept of dosage, as something that they considered when questioning students. Teacher 1 explicitly highlighted that she liked the idea of dosage because she interpreted as meaning she could make incremental changes to her instruction as opposed to altering the lessons dramatically. Teacher 1 also made the explicit connection between the concept of dosage and the *interactive* mode of ICAP, suggesting that she was trying to increase the dosage of questions and activities that elicit the two higher modes of engagement. Teacher 2 mentioned dosage in passing but seemed to be sharing that her goal was to add in "doses of good questions" when possible and that it was easiest when responding to students' answers.

These interviews provide us with some indication that teachers were mindful of the professional development while they were teaching.

Discussion

Changes in teacher questioning

This study investigated the effect of a theory-driven professional development module on middle school science teachers' questioning in their classrooms. To investigate the impact of the PD we articulated a coding scheme based on the ICAP framework to categorize teacher questions based on their accordance with the four modes of cognitive engagement theorized by ICAP. The findings here serve as preliminary evidence that a brief PD based on ICAP and utilizing best practices from the literature on teacher training, increases the number of constructive questions that teachers use in their teaching. Following the PD, the participating teachers increased their use of constructive questions and provision of interactive time in comparison to their own instruction prior to the PD. They also increased the number of constructive questions used in their follow-up questions. Given our small sample size, we can only argue that the changes are most likely due to the PD for a number of reasons.

First, the similarity between the question frequencies in the baseline lessons of both 7th grade teachers suggests that the measurements were representative of the types of questions typically asked at this level of science instruction. This finding is supported by many other studies of teacher questioning showing that teachers do not ask many deep questions (Dillon, 1988, Graesser & Person, 1994). Moreover, the overall pattern of these business-as-usual classes is in line with a number of studies that reported that teachers asked primarily low-level (e.g., *active*) questions during science instruction (Benedict-Chambers et al., 2017; Eshach et al., 2014). Taken together, these points suggest that we can assume with some certainty that during typical instruction, regardless of the lesson, these teachers would rely on *passive* and *active* questions without including *constructive* questions in their instruction.

A second piece of evidence for why the PD is the likely agent of change in the teachers' instructions is that the pattern of lecturing and how questions were embedded in the lectures, did not change for the teachers between their pre-PD and the post-PD lessons. Teacher 1 demonstrated a lecture-then-assess pattern for all three observed lessons, while Teacher 2 demonstrated a more rapid-fire style of questioning throughout the lectures. This consistency makes sense given one of the points emphasized in the PD was that teachers did not need to make wholesale changes to their instructional practice, but instead they should strive to increase the students' opportunities to enact constructive and interactive processes. The predicted change based on this aspect of the PD would be precisely what we saw in our results, a greater number of constructive questions and interactive opportunities embedded in the pacing and interactional patterns with which the teacher was already comfortable.

Finally, the coding of questions demonstrates that the teachers were sensitive to each lesson's prioritized mode of engagement. Teachers were asked to teach one lesson more interactively and one lesson more constructively and their questioning suggests that they increased the number of constructive questions for the constructive lesson and incorporated more interactive time during the interactive lessons. The lesson plans as provided by the research team were not prescriptive and so these differences in lesson questioning are most likely due to teachers' understanding about eliciting *constructive* and *interactive* processes.

Student learning gains

Our preliminary findings also show that students demonstrate learning gains in classes with a high number of constructive questions. Given that the students only received one lesson on two topics that were not part of their 7th grade curriculum, the percentage increase is remarkable, however, without a control classroom the magnitude of the effect is still in question. It is also noteworthy that the classes that had more opportunities to interact with one

Table 4. Student learning gains for both lessons.

Lesson	Unit		
Create-a-monster	Genetics		
	Gain scores M	SD	% Gain
Class 1 (constructive)	2.25	2.10	16
Class 2 (interactive)	3.60	1.50	25
Both	2.95	1.80	21
Lesson	Unit		
Pocket mice	Natural selection		
	Gain scores M	SD	% Gain
Class 1 (interactive)	2.41	2.00	20
Class 2 (constructive)	1.11	1.90	9
Both	1.76	1.95	14

another showed stronger learning gains, in line with predictions made by the ICAP framework.

Limitations

Due to the small sample size of students we have been careful not to make any far- reaching claims about the generalizability of student learning gains. We included the section on learning gains primarily as a small contribution to what we see as a large gap in the research literature on questioning practices in the classroom. Mainly, tying student learning to changes in questioning. Admittedly, with our data here, we can only conjecture about how the learning gains are related to the changes in questioning practice.

In this study, the teachers were working with fewer students than in their regular classrooms but we are not aware of any research that suggests a smaller class size would improve a teacher's questioning practice. Of course, we can imagine how a smaller class would facilitate certain instructional frameworks but we did not see anything in the literature we reviewed on questioning to suggest that teachers' reliance on primarily *closed* or *known answer* questions was driven by class size. Notably, the difference in class size was not mentioned by either teacher as a factor in their inclusion of more constructive questions, or in their use of more *active* questions in their instruction prior to the PD.

The participating teachers provided a recording of a lesson taught prior to the PD, that served as the baseline in our analysis. We have discussed the extent to which the questioning in these baseline lessons differed from the lessons the teachers taught following the PD. Because our professional development focused on how to elicit constructive and interactive engagement via questioning and activity prompts, we have argued that the changes in teacher questioning were likely brought about because of our PD. The extent to which other factors (e.g., providing the lessons' outlines) contributed to changes in teacher questioning is something we plan to explore in a larger study. With a greater number of participating teachers, we will be better able to the explore the nature and persistence of change in questioning due to our professional development.

Conclusion

This study represents our efforts to pilot a professional development module to support teachers' use of high-level questioning. We have shown evidence of the efficacy of ICAP as a framework for helping our participating teachers interject more constructive questions into their teaching. Anecdotal evidence also suggests that pairing ICAP theory with the concept of "dosage" may have facilitated these teachers' adoption of the theory as a means of monitoring and bettering the curricula they already have in place. This PD outlined here is intended to present a type of training module that serves to clarify a common imperative (e.g., active learning) and provides concrete markers of cognitive engagement from which teachers can base their decisions. These characteristics of the ICAP theory could improve teacher buy-in and facilitate the adoption of the theory into pre- existing teacher practice. Because of its ability to be interleaved with certain aspects of current practice, this PD could be implemented in schools where major revisions to teacher pedagogy and classroom communities are unsustainable.

Similarly, our coding scheme provides a tool for giving teachers feedback on their instruction. Because it makes fewer distinctions among constructive question types than other typologies, and focuses instead on accounting for the material presented to students, the coding scheme can be used while observing classroom instruction; allowing for timely feedback and concrete suggestions for teachers who want to increase cognitive engagement.

Acknowledgements

The opinions expressed are those of the authors and do not represent views of the Institute. The authors are pleased to acknowledge the contributions of Joshua Adams, David Yaghmourian, Natalie Newton, Nicole Boucher and Jacklyn Ho.

Funding

The research reported in this paper was supported by the Institute of Education Sciences, through the grant, *Developing and revising instructional activities to optimize cognitive engagement* (R305A50432) at Arizona State University, Michelene T. H. Chi, Principal Investigator.

ORCID

Joshua Morris http://orcid.org/0000-0001-7301-7485

References

Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. *International Journal of Science Education*, *22*(7), 665–701. doi:10.1080/095006900050044044

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of* *science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, *96*(4), 385–414. doi:10.1086/461835

Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1997). *Questioning the author: An approach to enhancing student engagement with text*. Newark, DE: International Reading Association.

Benedict-Chambers, A., Kademian, S. M., Davis, A. E., & Palincsar, A. S. (2017). Guiding students towards sensemaking: Teacher questions focused on integrating scientific practices with science content. *International Journal of Science Education*, *39*(15), 1977–2001. doi:10.1080/09500693.2017.1366674

Borko, H. (2004). Professional development and teaching learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Brickhouse, N. W. (1990). Teachers' beliefs about the nature of science and their relationship to classroom practice. *Journal of Teacher Education*, *41*(3), 53–62. doi:10.1177/002248719004100307

Chen, Y., Hand, B., & Norton-Meier, L. (2017). Teacher roles of questioning in early elementary science classrooms: A framework promoting student cognitive complexities in argumentation. *Research in Science Education*, *47*(2), 373–405.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, *6*(3), 271–315.

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., … Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*(6), 1777–1832. doi:10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. doi:10.1080/00461520.2014.965823

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking. *Journal of Research in Science Teaching*, *44*(6), 815–843. doi:10.1002/tea.20171

Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research*, *76*(4), 607–651. doi:10.3102/00346543076004607

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. doi:10.3102/0013189X08331140

Dillon, J. T. (1988). *Questioning and teaching: A manual of practice*. New York: Teacher's College Press.

Eshach, H., Dor-Ziderman, Y., & Yefroimsky, Y. (2014). Question asking in the science classroom: Teacher attitudes and practices. *Journal of Science Education and Technology*, *23*(1), 67–81. doi:10.1007/s10956-013-9451-y

Garet, M. S., Porter, A. C., Desimone, L., Birman, B., & Yoon, K. S. (2004). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*(1), 104–137. doi:10.3102/00028312031001104

Hargreaves, D. H. (1984). Teachers' Questions: Open, Closed and Half-Open. *Educational Research*, *26*, 46–51.

Hashweh, M. Z. (1996). Effects of science teachers' epistemological beliefs in teaching. *Journal of Research in Science Teaching*, *33*(1), 47–63. doi:10.1002/(SICI)1098-2736(199601)33:1<47::AID-TEA3>3.0.CO;2-P

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

Lee, Y., Kinzie, M. B. & Whitaker, J. V. (2012). Impact of online support for teachers' open-ended questioning in pre-k science. *Teaching and Teacher Education*, *28*(4), 568–577.

Masalmeh, J. (1998). *Effects of teachers' epistemological beliefs on their pedagogical content knowledge* (Unpublished master's thesis). Birzeit University, Birzeit, Palestine.

Michaels, S., & O'Connor, C. (2015). Conceptualizing talk moves as tools: Professional development approaches for academically productive discussion. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through talk and dialogue* (pp. 347–362). Washington, D.C.: AERA.

Monk, M., & Osborne, J. ed. (2000). *Good practice in science teaching: What research has to say*. Maidenhead: Open University Press.

National Research Council. (2012). *A framework for K-12 science education*. Washington, D.C.: The National Academies Press.

Oliveira, A. W. (2010). Improving teacher questioning in science inquiry discussions through professional development. *Journal of Research in Science Teaching*, *47*(4), 422–453. doi:10.1002/tea.20345

Resnick, L. B., Michaels, S., & O'Connor, M. C. (2010). How (well-structured) talk builds the mind. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in educational psychology* (pp. 163–195). New York, NY: Springer Publishing Company.

Tobin, K., & McRobbie, C. J. (1996). Cultural myths as constraints to the enacted science curriculum. *Science Education*, *80*(2), 223–241. doi:10.1002/(SICI)1098-237X(199604)80:2<223::AID-SCE6>3.3.CO;2-I

Van Zee, E. H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversation about science. *Journal of Research in Science Teaching*, *38*(2), 159–190. doi:10.1002/1098-2736(200102)38:2<159::AID-TEA1002>3.0.CO;2-J

Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *Journal of the Learning Sciences*, *6*(2), 227–269. doi:10.1207/s15327809jls0602_3

Windschitl, M. A., & Stroupe, D. (2017). The three-story challenge: Implications of the next generation science standards for teacher preparation. *Journal of Teacher Education*, *68*(3), 251–262. doi:10.1177/0022487117696278

Winne, P. H. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research*, *49*(1), 13–49. doi:10.3102/00346543049001013