Expanding Higher-Order Thinking Skills Through Academically Productive Talk in the Middle School Science Classroom

Jennifer Winsor, CAGS

New England College

For comments or questions for the author, contact Jennifer Winsor at jwinsor_gps@nec.edu

Abstract

In this small, mixed methods pilot study twenty-three 5th graders and two 6th graders from a large, urban middle school in the Northeast United States were given Accountable Talk (AT) sentence stems to use during a series of academically productive talk (APT) in their science classrooms over a 2-week period in the Spring of 2024. The goal of the study was to assess whether and how ATP, the independent variable, might lead to higher-order thinking (HOT), the dependent variable, an area of great focus within our school district and something which I believe is essential for success in life. I observed and recorded student discussions, and later evaluated the data using a higher-order thinking coding tool. Students were also given a CER-type (claim, evidence, reasoning) assessment at the conclusion of the intervention, which was compared against a similar assessment given to them in October 2023. The observed participants and two participating teachers were interviewed at the end of the study to triangulate quantitative results. While the study yielded modest quantitative results, a connection was made between the quality of the questions asked by the teachers and students' use of HOT in their discussions and their CERs. The 6th grade students demonstrated higher rates of HOT versus the 5th grade students, who had received less robust discussion questions, though the 5th students made more progress during the study than did the 6th grade students. From the results of this pilot study, I have recommended continued research into the important connection between student academically productive talk and higher-order thinking in our school.

Keywords: accountable talk (AT), higher-order thinking (HOT), academically productive talk (APT), claim, evidence, reasoning (CER) thinking, academic conversations

From my experiences as a teacher, I have come to believe academically productive talk (APT) and higher-order thinking (HOT) are both essential for student academic success. As a K–12 educator for 20+ years, I am passionate about student growth and achievement and eager to learn about effective strategies in the classroom. I am now an instructional coach at a Massachusetts public middle school, that has begun to emphasize the use of APT and HOT. However, our school and district have not yet reviewed the research or validity behind the tools we are investigating in regards to positive academic outcomes. Through researching student discourse and HOT as part of this study, I now appreciate just how critical APT and HOT really are. Prior research studies have indicated APT leads to better academic outcomes for students (Applebee et al., 2003; Aranda et al., 2020; Michaels & O'Connor, 2012; Osborne, 2010; Soysal, 2021; Soysal & Soysal, 2022). Further, there are multiple research studies showing APT *leads* to HOT in students (Perra et al., 2016; Soysal, 2021; Soysal & Soysal, 2022).

I believe students in the school district in which I work would benefit from educators that more deeply understand the value of APT and HOT skills and *how* to use and develop both in the classroom. However, the impact of APT and HOT go far beyond our city's public schools. All students deserve to experience engaging, challenging classes every day. Teachers who have their students argue compelling academic questions would certainly elevate the rigor of any classroom and instill higher-order thinking skills in young people (Perra et al., 2016; Soysal, 2021).

In this experimental design, mixed methods pilot study, students in 5th and 6th grade science classes at a large, urban middle school in the Northeast United States engaged for one to two weeks in purposeful, academically productive talk using an accountable talk protocol with an emphasis on building, sharing, and responding to arguments about topics aligned to a current science standard. Students discussed compelling questions that lent themselves to lively interactions and HOT. I observed and recorded a few students in each class multiple times over the course of the two weeks and coded student utterances for instances of HOT. At the end of the intervention, students completed a claim, evidence, reasoning written response (CER) on the same topic they discussed. This data was compared against a preintervention CER from earlier in the school year, which had already been graded. Following the intervention, I also interviewed the same students I observed to learn what the experience of engaging in regular classroom APT was like for them. I also interviewed the two teachers whose classes I observed to hear their thoughts on the impacts of the intervention in their classes. All data were used to triangulate my research questions around whether APT leads to increased HOT (verbally and in writing) in middle school science students. Research questions for this study included: 1. In what ways does academically productive talk (APT) in the science classroom relate to higher-order thinking (HOT) for middle school students in an urban school environment? 2. In what ways do compelling, higher order thinking questions, in the science classroom, relate to the quantity and quality of higher order thinking (HOT) for middle school students in an urban school environment?

Important Terminology

Academically Productive Talk (APT)

Academically productive talk is a process in which teachers encourage and support all students to think deeply and critically, articulate their reasoning, and listen with purpose in order to more deeply understand and integrate their learning. This is accomplished by structuring discussions that help students to articulate their thinking, exposing students to different points of view, asking students to respond to questions or challenges to their ideas, and revealing areas where more investigation might be necessary (Michaels & O'Connor, 2012; Shwartz et al., 2009).

Accountable Talk (AT)

Accountable talk (a form of ATP) refers to classroom discussions that are scaffolded by the teacher to allow students to engage in focused, coherent, and deep conceptual understanding with each other. Students are motivated to participate and eager to share their thinking with others. AT includes accountability to the community by showing respect and listening to students' peers, accountability to accurate knowledge—sharing evidence-based ideas, and accountability to rigorous thinking or the ability to accurately create and defend argumentative claims (Michaels et al., 2013).

Higher-Order Thinking (HOT)

Higher order thinking (HOT) goes beyond rote memorizing and the retelling of facts. HOT requires students to take what they have learned and do something with it, such as deeply analyzing the concept, connecting or applying it to a new situation or idea, evaluating its reliability or creating a new solution. According to the Hess Cognitive Rigor Matrix (CRM), the highest form of HOT in the math-sciences is the ability to synthesize information across multiple sources and/or to create or design a model to solve a practical or abstract situation (Hess, 2013).

Claim, Evidence, Reasoning (CER)

Claim, evidence, and reasoning (CER) refer to creating an accurate *claim* (i.e., an argument or a position on a topic based on sound *evidence*). Further, students must show the connection between their claim and evidence with *reasoning*. The ability to do these three steps is an example of higher order thinking as defined by Hess (2913)—to analyze and draw conclusions from data, citing evidence.

Accountable Talk Stems

Accountable Talk Stems are sentence stems or frames teachers frequently provide to students as they begin to engage in AT conversations. Common stems might include, “I agree with you because . . .” “I disagree because . . .” “I used to think this but now think that because . . .” These stems or frames can be helpful in holding students accountable to the core purposes of accountable talk when they are first starting out, but can be discontinued once students use them with fidelity on their own (Michaels & O'Connor, 2012).

Compelling Higher Order Questions

Compelling HOT questions are those that elicit higher order thinking and engagement from students. They are usually open-ended questions with no one right or wrong answer. Questions on controversial, ethical, and/or debatable topics are also compelling, students want to answer them. Some examples of compelling HOT questions include: (a) Do you think people in a poor community in India were justified in stealing fresh water from the neighboring wealthy community when they had no access to clean water? (b) Could scientists bring dinosaurs back to life and if so, how would they do this, and if they could, should they? These questions require students to apply their background knowledge, build a claim using evidence and reasoning, and potentially to engage in a back-and-forth APT discussion with their peers who may have a different position (Stanley, 2020).

Literature Review

According to The New Teacher Project (TNTP, 2018), the majority of U.S. students do not currently feel engaged or excited by their studies in grades 6–12 classrooms. The authors of TNTP (2018) reported classrooms are lacking regular student-centered activities such as discussions and arguments and truly cognitively demanding work that leads to HOT. Numerous studies, including major meta-analyses, have demonstrated the importance of classroom discourse around compelling questions and rigorous work as having a significant positive impact on student academic performance (Hattie, 2023). These and other studies have shown a connection between academically productive talk and higher-order thinking (Perra et al, 2016; Soysal, 2021; Soysal & Soysal, 2022). Science education, in particular, has been taught in a manner whereby the teacher knows all and gives unequivocal content to students (Osborne, 2010). Fortunately, this is changing with the adoption of Next Generation Science Standards in schools throughout the United States, including in Massachusetts [Massachusetts Department of Elementary and Secondary Education (MA DESE), 2016]. The present emphasis in science education has students doing the work of scientists and learning through inquiry and discussion.

This is encouraging, as much prior and current research ties student success in science to academically productive talk and work that elicits higher-order thinking (Osborne, 2010; Soysal, 2021). Another tangential benefit of APT is it tends to improve the culture of a classroom (Howell et al., 2011).

The Urgency of Higher-Order Thinking Skills (HOT)

According to the authors of this groundbreaking, massive study on student academic achievement in the United States (TNTP, 2018), only 17% of students met grade-level standards, while successfully mastering 71% of their schoolwork. The discrepancy lies in the quality of *content* and *instruction* many students, particularly students of color, those learning English (multilingual learners), and low-income students, receive day-to-day in school (TNTP, 2018). Their mixed-methods study was conducted in a variety of different school districts around the country representing a wide-mix of student demographics from elite charter schools to failing schools on the verge of being shut down. The researchers reviewed student data (grades 3–12), observed thousands of classrooms, and had students evaluate their own level of engagement throughout the school day. They also interviewed and surveyed students and teachers. TNTP (2018) found the key indicators of a high-quality academic experience included: (a) consistent opportunities to work on grade-appropriate assignments, (b) strong-instruction that allows students to do most of the thinking, (c) a sense of deep engagement in what students are learning, and (d) teachers who hold high expectations for students.

Across all classrooms TNTP (2018) observed, teachers were only offering students grade-appropriate work 26% of the time. Further, teachers on average allowed students to do the cognitive heavy-lifting in class for about 29 out of every 180 hours of core academic instruction. Additionally, half of the 2,000 students, who rated their level of engagement throughout the school day, experienced something in class that was both engaging and worthwhile. The researchers also reported, that number decreased as students went up in grade (TNTP, 2018). The long-term impacts of this lack of rigor and engagement effected more than low test scores. I believe students' livelihoods and futures are at stake; to me, this research indicated students simply cannot wait for engaging, challenging instruction and content.

The TNTP (2018) team laid out underlying reasons for these shortcomings in grade-level appropriate work and cognitive rigor. These issues are nationwide, not just in my school district or state. However, the bright side to this study, and the hope for CAMS, is the fact that educators can make shifts in their assignments, instruction, and engagement strategies, and increase their expectations of students. From the evidence in this literature review, I have developed a recommendation for further development of the system in our school to attempt to close the achievement gap by using student academically productive talk (APT) to engender HOT.

A focus on developing students' critical thinking skills is evidence-based as noted by Stanley (2020) in, *Promoting Rigor Through Higher Level Questioning*. Stanley wrote, "Using higher level questions in your classroom improves student achievement, builds understanding and retention of learning, increases student engagement, asks students to think for themselves, and teaches valuable 21st-century skills" (p. 13). Stanley also cited John Hattie's (2008) meta-analysis work that found a strong connection between student academic achievement and cognitively demanding work in the classroom. For example, cognitive task analysis, classroom discussions, evaluations and reflections, problem-solving teaching, and questioning all have an effect size well above .40, which makes them visibly significant. Hattie termed any effect size above .40 as being visible to the teacher in the classroom. Like TNTP (2023), Stanley (2020) wrote how little teachers engaged students in HOT and HOT questions in the classroom.

What Is Lacking in Science Education Today

Many researchers have written about the need for APT and HOT within science education. For example, Osborne (2010) argued science education was lacking in *argumentative discourse*. According to Osborne, "Critique is not, therefore, some peripheral feature of science, but rather it is core to its practice, and without argument and evaluation, the construction of reliable knowledge would be impossible" (p. 464). I believe, in conjunction with Osborne, students need to do the work of scientists and question and argue viewpoints, rather than learn a list of facts. Osborne noted several research studies showing students who participated in discourse or debate about science topics first, performed better on assessments of these topics. They noted the work of coming up with a logical argument called upon higher-order thinking skills.

Osborne (2010) wrote, "Argumentation in science education requires students to construct and evaluate scientific arguments to reason scientifically" (p. 465). Osborne's biggest point was argumentation is important in science classrooms and that it leads to HOT. Osbourne also talked about how science is often taught as a "monolith of facts" (p. 464), and this is not ideal or close to the real work in which scientists engage. They noted good *discussions* in the science classroom did not happen organically; they needed to be carefully structured by teachers. They ended by suggesting this is an area ripe for further investigation.

Many authors also described how classroom discussions are often one-sided and fall short of building HOT. Applebee et al. (2003), Michaels and O'Connor (2012), Shwartz et al., (2009), and Soysal (2021) all noted how the typical IRE format where a teacher initiates (I) a question to the entire classroom, a student responds (R), and the teacher evaluates (E) their response as to whether it was correct or not, does not lead to a rich discussion or higher-order thinking. Soysal (2021) explained when teachers want and seek out predetermined answers and establish single answers through evaluating and knowledge providing moves, there are fewer opportunities for students to make authentic intellectual contributions to classroom discussions. This IRE format happens all too often in many classrooms, including in science classes. According to Shwartz et al. (2009), science is social and *discussions* are one of the best ways to help students learn and construct understanding. Michaels and O'Connor (2012) wrote about how *accountable talk*, a form of APT, can and should be interwoven into the science classroom. They noted the use of evidence-based argumentation is important, such as the ability to play with ideas and change them when new evidence is presented.

HOT and APT discussions are also advocated for by the Massachusetts Department of Elementary and Secondary Education. The 2016 *Massachusetts Curriculum Frameworks for Science and Technology/Engineering Grades Pre-Kindergarten to 12* include nine guiding principles underlying all strong science content and instruction. Four of the Massachusetts science principles align directly with the Next Generation Science Standards. These include:

1. Guiding Principle 1 (Relevance): An effective science and technology/engineering program develops students' ability to apply their knowledge and skills to analyze and explain the world around them.
2. Guiding Principle 3 (Rigor): experimentation, design, and analytical problem solving are central to an effective science and technology/engineering program.
3. Guiding Principle 4 (Rigor): An effective science and technology/engineering program provides opportunities for students to collaborate in scientific and technological endeavors and communicate their ideas.
4. Guiding Principle 5 (Rigor): An effective science and technology/engineering program conveys high academic expectations for all students. (National Research Council, 2015, pp. 14–17)

The Next Generation Science Standards shifted science education from learning a series of facts from textbooks to hands-on, inquiry-based learning by students with a big emphasis on APT (National Research Council, 2015).

The Benefits of Academically Productive Talk (APT) and Higher-Order Thinking (HOT)

Many articles and other sources have gone into great detail about the benefits of APT and HOT. Some of those benefits have already been noted. Every research study I reviewed found a positive association between APT and student achievement (Applebee et al., 2003; Aranda et al., 2020; Osborne, 2010; Soysal, 2021; Soysal and Soysal, 2022).

Michaels and O'Connor (2012) summarized the myriad of benefits from APT. Specifically, APT provided a window into student thinking, boosted memory, led to richer associations, supported language development, encouraged deeper reasoning, including reasoning with evidence, and supported students' social development and social skills (Michaels & O'Connor, 2012). Michaels and O'Connor noted some of the unique features to science talk: (a) science discussions are focused on generating community-validated explanations of the natural world, using data and models as evidence or tools in developing ideas; (b) logical reasoning is central to scientific thinking; any credible theory must be concerned about and deal with contradictory evidence. Scientists must be willing to change their minds; (c) although scientists revolve their work in the specifics of their data, the ultimate goal is to generalize and create expanded explanations or theories.

Howell et al. (2011) conducted a research study where they interviewed middle school students about what they liked and did not like about academic talk in the classroom. Howell et al. discussed using accountable talk (AT) in an 8th grade social studies classroom. Howell et al. concluded:

> Findings indicated that students reported a better understanding of the content presented, a strong sense of community in the classroom, and an opportunity to develop an appreciation for the diversity of classmates' backgrounds, experiences, knowledge, and beliefs leading to a deeper, broader look at the social studies content presented. (p. 47)

Perra et al. (2016) authored a chapter in a book on best practices for bilingual learners (MLs). Their chapter on dialogic reasoning (DR), a form of APT, discussed the language development benefits of DR, particularly for MLs. Like Howell et al. (2011) article, Perra et al. noted how APT can also create a community of learners and improve relationships among students.

Connecting APT, HOT, & the Importance of Compelling Questions

Most of the articles and other sources identified in this literature review made a connection between APT and HOT. Articles defined HOT in a variety of ways. Michaels and O'Connor (2012) defined APT as:

> Everyone can hear and understand what is being said, so that every single student is part of the conversation. The conversation is focused, coherent, rigorous, and leads to deep conceptual understanding. Students are motivated to participate and want to go public with their thinking, feeling like they have a stake in the conversation. Conversation is not just for good talkers; everyone has a right and responsibility to contribute. The teacher guides students in practicing new ways of talking, reasoning, and collaborating with one another. (p. 1)

Perra et al., (2016), wrote, "Talk is a key social and cognitive tool that shapes reasoning and meaning making during different literacy practices" (p. 121). The authors use the term "interthinking" (Mercer et al., 2000, as cited in Perra et al., 2016) in regard to dialogic reasoning and HOT development. This same chapter also noted the importance of having a "big question" (Perra et al., 2016, p. 127) to drive the discussion, an important point in a few of the pieces I read and one I included in my research study.

Soysal (2021) conducted a few more recent studies in Turkey on APT and HOT. In their 2021 article, they developed an explanation about the importance of higher-order thinking in the science classroom. They noted such core components of critical thinking as: decision-making, inference, advanced clarification, analysis, evaluation, and explanation (and these are intertwined with scientific inquiry). This study concluded legitimating, challenging, monitoring, and evidencing are required to scaffold HOT. Like Osborne (2010), Soysal noted the importance of questioning, judging, or *arguing* in science discussions. Further, authors have noted argument construction, analysis, and evaluation are fundamental elements of critical thinking (Facione 1990; Golding, 2011; Siegel, 1988, as cited in Soysal, 2021).

Shwartz et al. (2009) explained teachers must pose good questions. Shwartz et al. stated, “Teachers should pose questions that push students to think more deeply about what they have observed, experienced, or read” (p. 45). Students need opportunities to question each other and think aloud with peers as they try to understand phenomena, etc. In this article the authors suggested brainstorming as a good way to make guesses (inquiry) at the start of a unit, from there, students talk with each other to come to a consensus on the cause of the phenomenon through a synthesizing discussion, and sensemaking is where students solidify their guesses and go deeper. They might challenge each other’s guesses with argumentation. Sensemaking may follow an experiment or investigation of some kind.

In Soysal and Soysal’s 2022 study, the authors noted open-ended questions elicited more HOT from students and led to better peer discussions. Like the other study by Soysal (2021), this one suggested the level of the teacher’s questions matters, and cognitively-demanding questions, such as requesting students to comment on others’ propositions yield HOT. In fact, student HOT was dependent on the quality of the teacher's questions. Eliciting questions encouraged the most student intellectual contributions to classroom discourse. When the teacher used peer-led evaluation questions (legitimating) they also got HOT from students. The study also mentioned that part of the legitimating process involved making students’ conceptual inconsistencies public.

Stanley’s (2020) book is all about the creation of higher-order thinking through the development of compelling questions. Compelling questions are open-ended and usually center around a highly charged topic, such as an ethical issue. I drew from this book while working with my science teachers to design questions for small group, classroom discussions and for the pre and post HOT assessments.

Methodology

This pilot study was a mixed-methods, experimental design research pilot study (Creswell & Guetterman, 2019) using a time series design approach. The use of the time series design prevented any threats to internal validity, such as selection errors, due to my not being able to randomly select student cohorts (science classes) to control or treatment groups. It also prevented threats to external validity, such as interaction of selection and treatment groups, as all students received the treatment. My participants included 23 5th graders and two 6th graders. This pilot study was created with the understanding this was the beginning of our process to study the use of academic productive talk (APT) and higher-order thinking (HOT) skills in our science classrooms.

My hypotheses for this study and included:

- Use of compelling, higher-order thinking questions in middle school science classrooms will lead to *longer* academically productive *discussions* among students.
- Use of an Accountable Talk (AT) protocol with an emphasis on building, stating, and responding to arguments will lead to *longer* APT among students in middle school science classrooms.
- Use of APT and compelling HOT questions in middle school science classrooms will lead to more *HOT utterances* during student discussions.
- Use of HOT questions and an AT protocol over an extended period of time (at least 2 weeks with use almost every time the class meets) will lead to *higher performance* (including more evidence of HOT) on a CER-type assessment of the topic or standard being taught (Posttest).

I also developed three secondary research questions for postintervention interviews with student focus groups and teachers that included:

- In what ways do compelling, HOT questions in the science classroom relate to HOT for middle school students in an urban school environment? (quantitative)
- What do students and teachers think of APT and HOT questions? (Qualitative, second portion of my study)

- In what ways does APT impact a classroom's culture? (Qualitative, second portion of my study)

My independent (treatment) variable was APT (academically productive talk with an emphasis on argumentation). The dependent variable (outcome) was the use of HOT while talking with peers (APT) and use of HOT as students answered a final assessment on the topic (in a claim-evidence-reasoning assessment).

As the treatment (APT, academically productive talk of compelling HOT questions) was used in classes. I observed and recorded two small groups of 5th graders (4 students in one class and 2 in another) and one duo of 6th graders a total of 7 times (3 for 5th and 4 for 6th). I then analyzed student discussions using a HOT Utterances coding tool I made from the Hess Cognitive Rigor Matrix for math and science (2013).

The 5th grade teacher had students discuss the following questions:

- On 4/30/24: Pick the most important paragraph in the reading on the lack of accessible fresh water for all people in the world and explain why. She then had students find and discuss the one sentence in the article that was most important and why.
- On 5/2/24 (two different groups recorded): Decide on 5–6 key problems related to your research project topic on how humans negatively impact the environment. One group focused on light pollution and the other on greenhouse gasses.

The 6th grade teacher had students discuss the following questions:

- On 5/6/24: If we could bring a 1 million old dinosaur back to life, should we and why?
- On 5/9/24: An organism has died next to a river. Describe one way it could become a fossil.

- On 5/10/24 (2 rounds): Looking at a diagram of fossils buried in different layers of earth: Which organism would make a good index fossil and why? (This discussion was run twice using different images.)

As soon as the research study ended, I conducted focus group interviews, with the students I observed, to hear directly from them what they thought of the APT, argumentation, HOT questions. I used open-ended questions from another study (Howell et al., 2011). This allowed me to triangulate my results and understand how my participants experienced the intervention. I also interviewed the two teachers, asking them the same questions. I was curious to learn if the use of AT improved the classroom culture, for example, as other research studies have shown this to be the case (Howell et al., 2011).

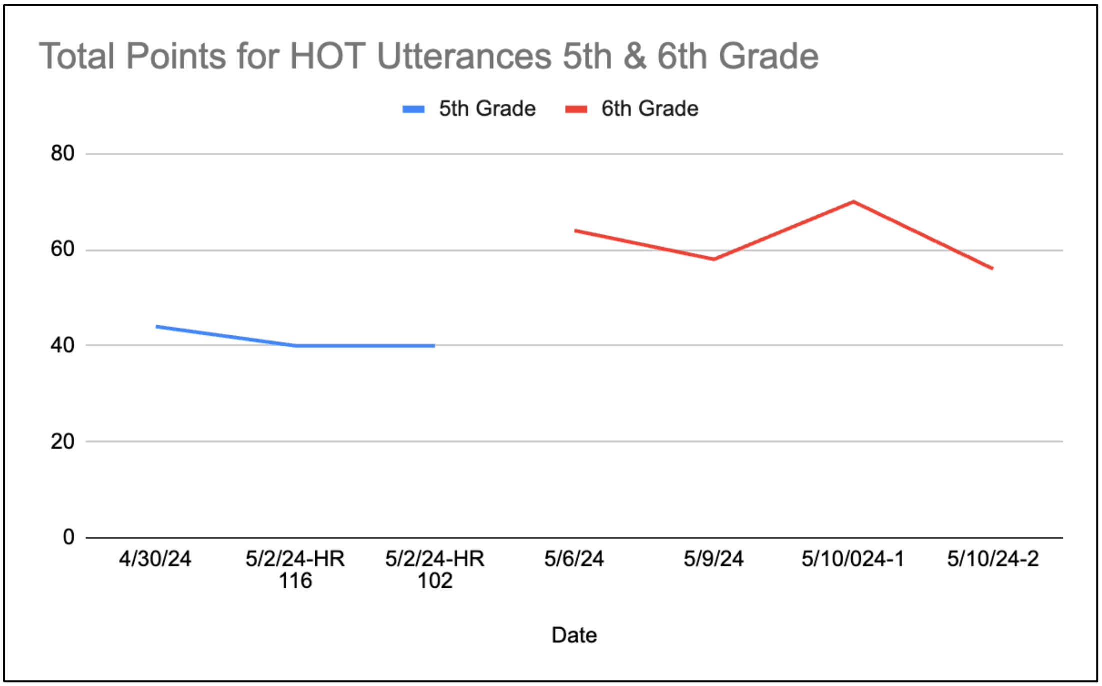
I assessed the impact of the intervention by comparing student results on a CER-type question on the topic and standard they had recently discussed against student CER data from earlier in the school year. This was graded using a district-created CER rubric, the same one used to grade student CERs earlier in the year.

Results

My HOT Utterances Coding Tool consisted of 20 rows for assessing different aspects of HOT during student small-group discussions (see Appendix A). The rows included categories to assess the length of the conversation, the use of the AT protocol sentence stems/frames, changes in thinking when new evidence was introduced, the use of academic language, and several categories specific to the Hess Cognitive Rigor Matrix for Math and Science (2013) such as, “Analyze and draw conclusions from data, citing evidence,” which is considered a Depth of Knowledge (DOK) 3 (Hess, 2013). Each row included a **rating** of 1–3 with 3 indicating the row was achieved/attained by all members of the group and 1 indicating it was done with less than 50% fidelity during the discussion and/or by less than 50% of the students in the group. For example, if students changed their thinking on a topic when new evidence was introduced by a peer, I would rate the row as a 3. Each row was also assessed by its **frequency,** 1–3, with 3 indicating it was observed 100% of the time by all students in the group and 1 indicating it was observed less than 50% of the time and/or by less than 50% of the students. I added the total rating and frequency scores together for each group I observed over the 2 weeks and entered the data into Google Sheets (see Figure 1).

Figure 1

Total HOT Points



I observed a few trends from the data in Figure 1. First, the 5th graders, whom I observed on 4/30 and 5/2, showed slightly *less* HOT in their discussions on the *second* observation (40 total points) than the first (44 total points). It is interesting the 5th grade students found the second question to be slightly more difficult to enter into more HOT conversations. From my observations of the 5th grade students, I concluded that fifth graders had more need of the sentence stems for their second topic. From their statements in the interview, the 5th graders appreciated the stems more than did the 6th graders.

I also observed the 6th graders engaged in more HOT across all observations than the 5th graders with an average score of 62 as compared to 5th grade average of 41.3. Sixth grade scores for each session included: 64 on 5/6, 58 on 5/9, 70 on 5/10-1 and 56 on 5/10-2, all substantially higher than those from the 5th grade. This is not surprising, as I would expect the older students, with 1 more year of experience to show more maturity in their discussions than the younger 5th grade students. Also, because the fifth grade had already developed a project to implement during this time, the teacher in grade five used other materials than those created for this project, and I am not sure how that affected the students' discussions. Although the 6th graders used more higher-order thinking in their discussions, the student scores fluctuated up and down, depending on the students and the ideas they were discussing. This leads me to believe the levels of questions and the discussion topics need to be developed with utmost care.

As part of the quantitative portion of my study, I also compared CER test scores (see Figure 2) from October 2023 and immediately after the AT intervention for the 5th and 6th graders. Their CERs were graded using our public school's Argumentative Writing Student Rubric (see Appendix B) in October and May. The rubric assesses students on the quality of their claim, evidence, and reasoning, and gives them a total aggregate score.

Figure 2 compares the CER sub-categories (claim, evidence, and reasoning) as well as the total scores between October 2023 (preintervention) and May 2024 (postintervention) for the 5th grade. The CER assesses student *argumentation*, which is considered an aspect of HOT. Figure 2 indicates, overall, the 5th graders *improved* in their writing and HOT in two (evidence and reasoning) of the three CER categories and obtained the same score in the third category (claim) between October 2023 and May 2024. The 5th grade students raised their total score by 15%.

Figure 2

Grade 5 CER

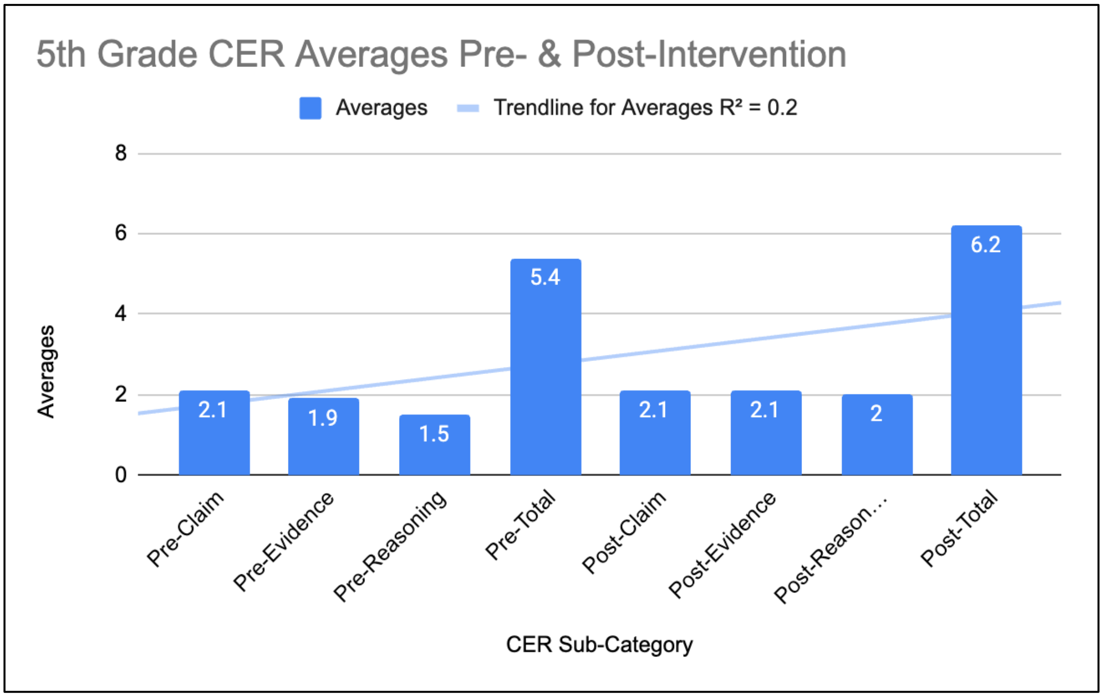
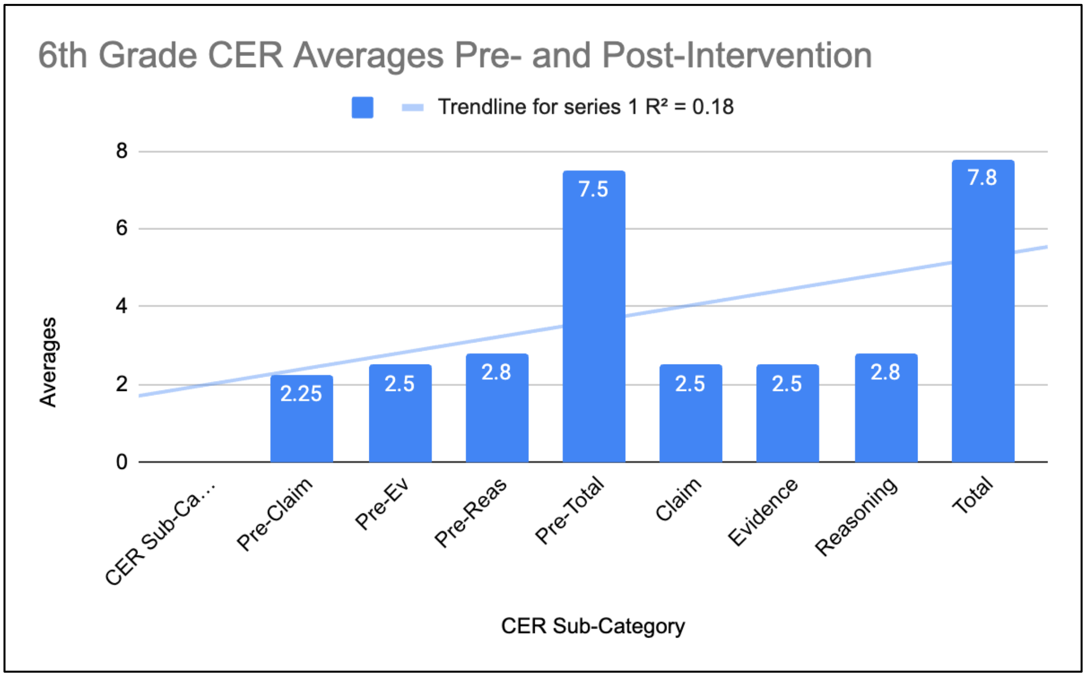


Figure 3 compares the CER sub-categories (claim, evidence, and reasoning) as well as the total scores between October 2023 (preintervention) and May 2024 (postintervention) for the 6th grade (see Appendix B). Like the 5th grade scores, the two 6th graders improved between the fall and spring, but not as dramatically. The 6th graders improved in the quality of their claims slightly, while achieving the same scores for evidence and reasoning. The 6th graders raised their total score by 4%.

Figure 3

Grade 6 CER



Figures 2 and 3 include a regression model trendline created through Google Sheets. The R-squared indicates the proportion of the difference in the dependent variable attributable to the independent variable(s) in a regression model. It measures the goodness of fit to the observed data, indicating how well the model's predictions match the actual data points (Investopedia, n.d.). The trendline goes up for both the 5th and 6th grade.

Qualitative Phase

At the end of the intervention all observed students (5 in 5^{th} and 2 in 6^{th}) and the two teachers were interviewed in regards to their thoughts about using AT in the science classroom. All students received the same questions adapted from Howell et al. (2011). The five prompts were also adapted and given to teachers. The five prompts for students included:

- Accountable Talk helped me understand science better.
- Accountable Talk helped me feel comfortable talking to my peers about science.

- Accountable Talk made me feel responsible for my learning in science.
- Accountable Talk helped me understand how to communicate better.
- Accountable Talk help me create a community in our classroom.

The five prompts given to teachers included:

- Accountable Talk helped my students understand science better.
- Accountable Talk helped my students feel comfortable talking to their peers about science.
- Accountable Talk helped my students take ownership of their learning in science.
- Accountable Talk helped my students understand how to communicate better.
- Accountable Talk helped my students create a community in our classroom.

Among the 5th graders there was wide agreement AT led to more and better discussions in class. In fact, students indicated they wish more of their teachers used the talking stems/sentence frames. One student stated, “We actually talked more than we usually do.” Another student said, “Using a sentence starter helped.” Another added, “It made me comfortable because I knew if I was using the sentence starter and I didn’t know what a claim would be, I could use that.” A final student stated, “We would talk more in our other classes if we used it.”

Students in the 5th grade also all strongly agreed with the statement AT helped them understand how to communicate better. One student said, “I could use the agreement or clarification if [I was] confused, if someone didn't hear or didn’t understand.” Another student described how AT helped create more community in their classroom stating, “It gets people talking to each other and they can become friends or disagree in a respectful way.”

The two 6th graders, on the other hand, did not believe the AT sentence frames/stems had helped in their conversations. One 6th grade student stated, “Talking stems is a way to start a sentence, but it didn’t help.” The second 6th grader added, “It didn’t affect much of my responsibility, because, I mean, like, it's just an option you can use to start sentences." The two sixth grade students also did not agree AT had improved their classroom community, which they saw as severely lacking. Nor did these students feel AT improved their overall communication skills. However, the two did see the value in student academic talk in general. One student said, "But the talking did help, we could understand each other's perspective when we talked to each other." The second sixth grade student remarked, "We got more intel, you get more information on one subject."

The two teachers both agreed on the theoretical value of the use of AT stems/sentence frames, but were not convinced they had impacted their recent classes very much. They were not sure that this project had enough time to have a full impact on their students. Some of their positive views on AT included: "When a student says something wrong and another student is able to correct them and explain why, the first student will have a better understanding of the material." The other teacher added, "I think it presents them with an opportunity to engage with what they already know and question what they don't, to see if they can get a better handle on the unknown." One teacher also stated, "Yes, the sentence stems gave the other kids something to start with [as opposed to the two I observed in 6th grade]." The second teacher agreed, stating, "Yes, it helps the kids learn better, gives them an opportunity to share and question each other about what they do and don't know. I can hear and adjust my practice too (based on what they share with each other)." Neither teacher thought the use of AT had contributed much to the sense of community in their rooms over the short cycle of use in the late spring. One teacher stated, "No, the sense of community is all within the first month of school and [created] outside the classroom." The second teacher had a similar feeling, "I think if you specifically taught this in the beginning of the year, it would have more of an impact, but starting it in April/May, less so." Finally, one teacher did point out the particular value to ML (multilingual learners) students, which the majority of students in our district are: "Yeah, they had something that showed them what's appropriate to say in the situation. And, a lot of them are former ELLs and that helps as well." Both teachers indicated wanting to increase how often they use AT in their classrooms and start in earnest at the beginning of next school year.

Discussion

The TNTP (2018) data, that describes the lack of HOT and of rigor in many schools, aligns with my experiences in education over the last 20+ years as a teacher, administrator, and instructional coach. I work at the Cherry (pseudonym) Avenue Middle School (CAMS) in a city in Massachusetts, serving grades 5–8. Our students' MCAS data is among the lowest in the state, putting us at the 2nd percentile for ELA and math. At the same time, our school has some of the highest-need students in the Commonwealth, with 84% low income, 28% MLs, and 18% with disabilities. According to our 2022 district's state report card (Massachusetts Department of Elementary and Secondary Education, n.d.)., the majority of our students are Hispanic or Latino (91%).

A major academic focus for administration in the Cherry Public Schools is on higher-order thinking. Teachers are rated on this during administrator observations using both the Hess Cognitive Rigor Matrix (Hess, 2013) and the SchoolWorks (2022) Classroom Visit Tool. Further, SchoolWorks, an outside agency, Cherry Public School leaders, and our internal team made up of Instructional Leadership Team members, regularly conduct instructional walkthroughs of the building. Unfortunately, the results thus far have been mostly unsatisfactory, with only a very small percentage of classrooms engaging 50% or more of their students in higher-order thinking as defined by Hess and SchoolWorks. I believe our teachers and students need more assistance and training to close the learning gaps for our students. Our teachers work hard for their students, and I believe that ATP and HOT can help our school move forward.

It was clear from the responses of the 5th grade students that they felt APT was very helpful in their discussions. The CER scores for the fifth-grade students were positive, with the students raising their scores in two of the three CER areas (i.e., evidence and reasoning) from the pre to the posttest, while maintaining their score in the claim category. These results are very promising, especially considering the project took place in the spring of the year, when there were many distractions with end-of-the-year activities. The gains made by the fifth graders were also impressive because they were accomplished in a short period of time. I intend to take what I have learned from this pilot study to expand my focus to more teachers in the building and to include my findings in future professional development sessions. This could allow me to engage in a full-length study of the school.

As noted, the 6th graders engaged in more HOT on all observations than the 5th graders. It could be in part due to the differences between prompts, and it is also likely due to the fact the 6th grade students are older and more experienced students. The two 6th graders I observed appeared to be academic leaders in the classroom. This is probably why neither of them needed nor saw much value in using the AT sentence frames (see Appendix C). From my observations, it is clear that different kinds of students need more specific ways to encounter the curriculum and the accompanying discussions. This is an important observation as I move from this pilot study to a full-length study for our school.

Limitations

There were five limitations to my study: (a) length of the study; (b) size of the participant pool; (c) lack of sufficient training for teachers or students, (d) timing of the study, and (e) inconsistencies in the assessment tool used in the study. The gains made in this pilot study were accomplished in spite of these major limitations for the study. With these limitations, I am excited to take the data and learning experiences to develop a full study for the coming school year.

First, this pilot study occurred over a two-week period. The students made gains in their HOT skills in a very short period of time. I did not have the time needed to help teachers learn how to create and implement multiple kinds of HOT questions. This also meant that students were not well prepared to employ ATP discussion techniques in their work. There was no time for teachers or students to learn and practice the skills required for effective ATP and HOT implementation. The lack of time also meant that I could not accomplish pre and poststudy observations of the teachers nor the students.

Second, my sample size was very small. There were six grade five students and two grade six students. I observed the students in three groups: one grade five group of two students, one grade five group of four students, and one grade six group of two students. Although, I developed some interesting data, the utility of the data is limited. Since this was a pilot study, I did gather some important ideas to use in the full study. I will work to ensure the participant pool is larger and more diverse in the full study. Since the student body in our school is diverse, this will be very important for me to accomplish.

Third, since the time span to create and implement this pilot study was limited, I was not able to create effective training for the teachers and the students. In order to gather more meaningful data, it will be important to make sure that teachers understand and have gained the skills necessary to create higher order thinking questions for their students. Teachers should also be trained to help their students learn how to apply ATP discussions in their work. Both teachers and students should be given time to practice so their discussions can be more productive, especially in terms of HOT skills. It is also important for teachers and students to gain a full understanding of the ATP process and practice using it regularly in their classroom.

Fourth, the timing of this pilot study was limiting in various ways. This study was implemented during the spring term of the school year. End-of-the year activities and preparations caused conflicts in time and the ability to work with students and teachers. Spring assessments also caused time conflicts in implementing this study. In the fifth-grade team, the teachers had created and were implementing a grade five project. That meant the fifth-grade teacher had to use the process and questions their team had already developed. Though the students were engaged in discussions for their project, the prompts given to the students were not created to specifically be HOT questions. During this time period, students tend to lose some of their concentration on school work. The end-of-the-year processes usually make teaching and learning become more complicated. In spite of those limitations, the grade five and grade six student participants raised their scores during the study. That result is very promising.

Fifth, due to the time constraints, I was not able to validate the observation tool used in this study. In the full study, I will make sure the tools used are valid and reliable. The CER assessments were scored by one group of teachers in the fall, and I assessed the CER in the spring for my participants. Thus, even though the same tool was used to score students, there could have been inconsistencies with different observational abilities and training between other teachers and my abilities and training. Thus, the scores cannot not be considered to be as valid in this pilot study.

In spite of the limitations of this pilot study, I am encouraged by the fact that all three groups of students made progress. That is an encouraging sign for the future and a full study, free from the current limitations. In the interviews students and teachers were encouraging me to continue to study these ideas and practices for ATP and HOT.

Future Research

While the results from this small pilot study were limited and not robust, I saw, and the students and teachers indicated the value of AT in the classroom. Further, students now clearly see the value of AT, APT, and AT sentence stems/frames, and some students are asking for them to be used more often by their teachers. The data from the study also demonstrated to me that the quality of the questions being asked by the teacher matter. This is an area our school addressed in our Annual Improvement Plan for the 2024–2025 school year, as it has also been seen by outside observers during Instructional Rounds. APT, when paired with compelling questions, can be very powerful, and I believe APT concepts should be explicitly taught to students at the start of the school year. My review of the data from students in this pilot study, along with the teacher interviews indicated connections between AT, ATP, and HOT. I have concluded it is critical all students develop HOT skills so they can succeed in any schooling or career in the future and become deep thinkers and engaged citizens.

These conclusions have led me to four recommendations for research in our school. My first recommendation is to develop AT stems for our sciences classes and gather data from teachers and students as to their effectiveness. Then, from that data our teachers could refine their AT stems to make sure they are more effective for our students. As we implement the stems with students, it will be important to gather data to assess which AT stems are most effective for our students. This is especially important for our fifth-grade students, who found the AT stems to be very helpful. I also recommend that we work to develop more complex stems for grade six students, who are more mature and experienced than their younger peers.

The second recommendation for our school is to participate in professional development in creating compelling questions for students to utilize in their APT activities. In this process, it seems likely to me that teachers will need multiple levels of compelling questions, since our students are diverse as learners. During such professional development sessions, teachers could develop compelling questions for their most important assignments and projects. Then it will be important to gather data to assess the compelling questions in order to know whether or not we will need to make adjustments for some questions.

My third recommendation is that future research on AT and HOT should involve a much larger sample size over an extended period of time in the school year. I am encouraged because our school is moving in the direction of developing APT and HOT and excited to continue this process. It will be important to develop research during this process that employs a large group of students. The results from the pilot study are encouraging, and it will be important to gather enough data to allow us to make more certain recommendations for our teachers and students.

My fourth recommendation is to develop research in relation our professional development activities. Any school will only make progress if teachers are properly trained and supported, the only way to understand how well professional development and support for teachers are working is to gather appropriate data to give our school the ability to assess our progress. I believe our school has the appropriate elements in place to accomplish the development of a valid research process. If our process is successful, then we would be able to share our results with other schools.

Final Thoughts

I am excited about the direction our school is taking in relation to AT, APT, and HOT. Our teachers are dedicated, and, if supported appropriately, will accept the challenges of closing the academic gaps for our students. These concepts have been shown to work elsewhere, and I am confident that we can develop similar results. Maybe even more important for me, if we are successful in this process, we will give our students skills that will help wherever they go after high school. We have the opportunity to help our students to actually be prepared for college and the world of work in their future lives. Our students will have gained the HOT skills and communication skills to be successful in their future endeavors. I believe this work will lead our students to become life-long learners and caring and positive citizens.

References

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, *40*(3), 685–730. https://doi.org/10.3102/00028312040003685

Aranda, M. L., Lie, R., & Selcen Guzey, S. (2020). Productive thinking in middle school science students' design conversations in a design-based engineering challenge. *International Journal of Technol Design Education*, *30*, 67–81. https://doi.org/10.1007/s10798-019-09498-5

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson Education, Inc.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The Delphi Report. California Academic Press. https://www.qcc.cuny.edu/socialSciences/ppecorino/CT-Expert-Report.pdf

Golding, C. (2011). Educating for critical thinking: thought-encouraging questions in a community of inquiry. *Higher Education Research & Development*, *30*(3), 357–370. https://doi.org/10.1080/07294360.2010.499144

Hattie, J. (2008). *Visible Learning*: *A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hattie, J. (2023). *Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement* (1st Ed.). Routledge.

Hess, K. (2013). *A guide for using Webb's depth of knowledge with common core state standards*. Common Core Institute. https://www.casciac.org/pdfs/Webbs-DOK-Flip-Chart.pdf

Howell, P. B., Thomas, S., & Ardasheva, Y. (2011). Talk in the classroom: Meeting the developmental, academic and cultural needs of middle school students. *Middle Grades Research Journal*, *Volume 6(1)*, 47–63. https://www.academia.edu/40339914

Investopedia. (n.d.). *R-Squared: Definition, calculation formula, uses, and limitations*. https://www.investopedia.com/terms/r/rsquared.asp#:~:text=R%2Dsquared%20tells%20you%20the,match%20the%20actual%20data%20points

Massachusetts Department of Elementary and Secondary Education. (n.d.) *School and district report card, 2022*. https://reportcards.doe.mass.edu/2022/00570050

Massachusetts Department of Elementary and Secondary Education. (2016). *Massachusetts curriculum framework 2016: Science and technology/engineering grades pre-kindergarten to 12*. https://www.doe.mass.edu/frameworks/scitech/2016-04.pdf

Michaels, S., & O'Connor, C. (2012). *Talk science primer.* TERC: An education research and development organization. https://inquiryproject.terc.edu/shared/pd/TalkScience_Primer.pdf

Michaels, S., O'Connor, M.C., Williams Hall, M., & Resnick, L.B. (2013). *Accountable talk sourcebook: For classroom conversations that work*. Institute for Learning, University of Pittsburgh. https://www.fredhutch.org/content/dam/www/about-us/education/sep/2020/AT-Sourcebook.pdf

National Research Council. (2015). *Guide to implementing the next generation science standards.* National Academies Press. http://www.nap.edu/catalog/18802/guide-to-implementing-the-next-generation-science-standards

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977) New Series, 463–466. American Association for the Advancement of Science. https://www.jstor.org/stable/40655777

Perra, M. O., Wagner, C. J., Procter, P., Leighton, C. M., Robertson, D. A., Paratore, J. R., & Ford-Conners, E. (2016). Chapter 7: Dialogic reasoning: Supporting emergent bilingual students' language and literacy development. In P. C. Proctor, A. Boardman, & E. H. Hiebert (Eds.), *Teaching emergent bilingual students: Flexible approaches in an era of new standards*, 119–137. The Guilford Press.

SchoolWorks. (2022). *Classroom observation manual (CVT)*. https://www.schoolworks.org/service-type/schoolworks-instructional-supports/

Shwartz, Y., Weizman, A., Fortus, D., Sutherland, L., Merrit, J., & Krajcik, J. (2009, Summer). Talking science: Classroom discussion and their role in inquiry-based learning environments. *The Science Teacher (National Science Teachers Association)*, *76*(5), 44–47. https://weizmannresearchmanagement.esploro.exlibrisgroup.com/esploro/outputs/journalArticle/Talkingscience/993289251203596?institution=972WIS_INST

Soysal, Y. (2021). Talking science, *Science & Education 30*, 33–65. https://doi.org/10.1007/s11191-020-00163-1

Soysal, Y., & Soysal, S. (2022). Exploring science teacher questions' influence on the students' talk productivity: A classroom discourse analysis approach. *SAGE Open*, *12*(2). https://doi.org/10.1177/21582440221102433

Stanley, T. (2020). *Promoting rigor through higher-level questioning: Practical strategies for developing students' critical thinking*. Prufrock Press Inc.

The New Teacher Project (TNTP). (2018). *The opportunity myth: What students can show us about how school is letting them down–and how to fix it*. The New Teacher Project, Reimagine Teaching. https://opportunitymyth.tntp.org/

Appendix A

Coding Tool for AT & HOT Utterances During APT Small-Group Discussions

Anticipated group size = 3–4 students *Anticipated discussion length = 5–15 minutes*

Date ______ Classroom Number _____ Student ID #s _____________________

Accountability to Discourse & Group

Indicator	Rating (1-3) *	Frequency (1-3) **	Notes
Everyone speaks and articulates loudly and clearly enough for all group members to hear and understand what is being said			
Students ask each other *clarifying questions* if they misunderstand or mishear a peer			
Everyone participates at least once during the discussion			
Students use AT sentence frames to *state* ideas/claims/opinions			
Students use AT sentence frames to *respond to* peer Ideas, claims, opinions (elaborate, provide contradictory information, etc.)			
Students use AT sentence frames to *ask* peers questions (beyond clarification)			
Students shift/change thinking, opinion, claim when new or *contradictory evidence* is presented by a peer			
Students use *academic language* specific to the topic			
Students *remain on-topic* (APT) the entire length of the time set by the teacher			
Discussion lasts *entire* length of time set by teacher (usually 5-15 minutes max)			

Hot Utterances

Indicator	**Rating (1-3) ***	**Frequency (1-3) ****	**Notes**
DOK 3/Understand Explain, present, generalize, or connect ideas *using supporting evidence* (Make a claim with *supporting evidence*)			
DOK 3/Analyze Analyze and draw *conclusions* from data, citing *evidence* (inference-making)			
DOK 3/Evaluate Evaluate validity and relevance of *evidence* used to develop an argument or support a perspective			
DOK 3/Evaluate Verify or critique the accuracy, logic, and reasonableness of stated conclusions or assumptions			
Include *reasoning* (the why) behind a claim and evidence			
DOK 3/Apply Apply a concept being discussed to a new context (transfer knowledge)			
DOK 3/Analyze Analyze or interpret *interrelationships* among concepts, issues, and/or problems (inference-making)			
DOK 3/Analyze *Compare* or *contrast* information within or across data sets or texts (or an idea proposed by a peer)			
DOK 3/Analyze Use *reasoning* and *evidence* to generate criteria for making and supporting an argument of judgment			
DOK 3/Create Develop an alternative solution or perspective to one proposed (e.g., debate)			

Scoring Rubric

*Rating Scale
3 = done with 100% fidelity by all students in the group 2 = done with 50% fidelity and/or by 50% of students in the group 1 = done with less than 50% fidelity and/or by less than 50% of the group
**Frequency
3 = observed 100% of the time by all students in the group within a 5-10-minute interval 2 = observed 50% of the time and/or by 50% of the students within a 5-10-minute interval 1 = observed less than 50% of the time and/or by less than 50% of the students within a 5-10-minute interval

Appendix B

Argumentative Writing Student Rubric

	Meets Expectations (3)	**Approaching** (2)	**Emerging** (1)*
Claim	Restates and answers the research question Accurate conclusion based on data collected or presented Complete sentence	Correctly answers the research question Accurate conclusion based on data collected or presented Phrase or complete sentence	Restates the Research question with misinformation or Answers the research question correctly with one or two words (e.g. "*Yes*." or "*Kinetic energy*.")
Evidence	Lab results and data are described thoroughly in bullet points or complete sentences. Data are sufficient, relevant, and support the claim. Refers to appropriate figures, graphs, texts, etc. to describe both qualitative and quantitative data, with appropriate units for quantitative data	Lab results and data are described in bullet points, phrases, or complete sentences. Data are relevant and support the claim. Includes qualitative and/or quantitative data	Lab results and data are included. Some data are relevant and support the claim.
Reasoning	Reasons are explained in an effective way that compels the reader to accept the claim, including *how* and *why* the data supports and defends the claim. Relevant scientific principles, facts, theories, phenomena, or concepts are described extensively, and key scientific vocabulary terms are included and used correctly. Uses suitable words such as conjunctions to connect sentences in a logical sequence.	The evidence is explained in phrases or complete sentences. Some science and key vocabulary terms are included and used correctly.	The evidence is restated. Some science and key vocabulary terms are included.

Appendix C

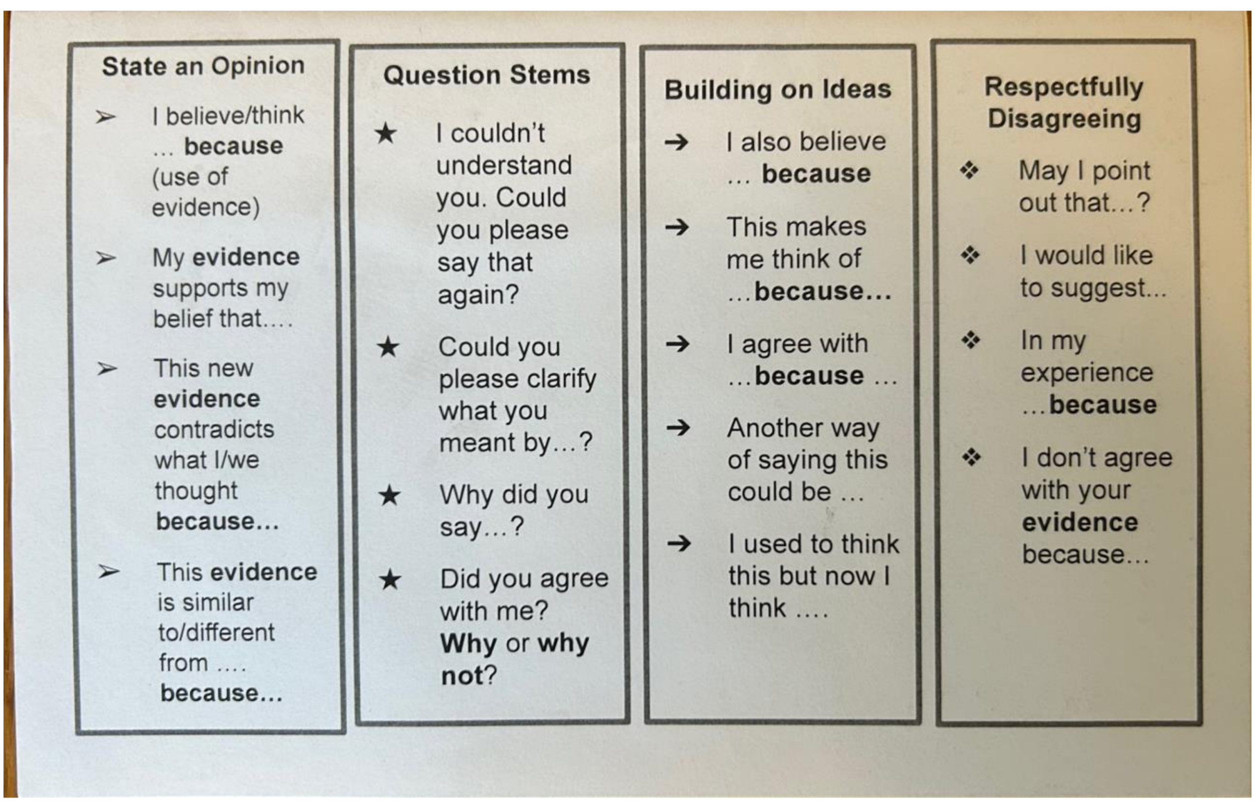
Accountable Talk Sentence Stems/Frames

Grade 5

ACCOUNTABLE TALK

Agreement I agree with what ___ said because… I agree with ____, but I would add….	Disagreement I disagree with what ____ said because… I am not sure I agree with what ____ said because…
Clarification Could you please repeat that for me? I am not sure I understood when you said ____, could you say more about that?	Confirmation I believe… I think… I found further evidence of what you said about…
Confusion I don't understand… I am confused about… I am not clear on…	Extension I was thinking about what you said and I was wondering what if… This makes me think…. I want to know more about…

Grade 6



State an Opinion	Question Stems	Building on Ideas	Respectfully Disagreeing
➢ I believe/think … **because** (use of evidence) ➢ My **evidence** supports my belief that…. ➢ This new **evidence** contradicts what I/we thought **because…** ➢ This **evidence** is similar to/different from …. **because…**	★ I couldn't understand you. Could you please say that again? ★ Could you please clarify what you meant by…? ★ Why did you say…? ★ Did you agree with me? **Why** or **why not**?	→ I also believe … **because** → This makes me think of …**because…** → I agree with …**because** … → Another way of saying this could be … → I used to think this but now I think ….	❖ May I point out that…? ❖ I would like to suggest… ❖ In my experience …**because** ❖ I don't agree with your **evidence** because…