International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

The Learning Benefits of Being Willing and Able to Engage in Scientific Argumentation

Meghan Bathgate[a], Amanda Crowell[b], Christian Schunn[a], Mac Cannady[c] & Rena Dorph[c]

[a] Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA, USA

[b] Eskolta School Research & Design, New York, NY, USA

[c] University of California, Berkeley, CA, USA

Published online: 26 May 2015.

To cite this article: Meghan Bathgate, Amanda Crowell, Christian Schunn, Mac Cannady & Rena Dorph (2015): The Learning Benefits of Being Willing and Able to Engage in Scientific Argumentation, International Journal of Science Education

To link to this article: http://dx.doi.org/10.1080/09500693.2015.1045958

International Journal of Science Education, 2015
http://dx.doi.org/10.1080/09500693.2015.1045958

The Learning Benefits of Being Willing and Able to Engage in Scientific Argumentation

Meghan Bathgate[a*], Amanda Crowell[b], Christian Schunn[a], Mac Cannady[c] and Rena Dorph[c]
[a]*Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA, USA;* [b]*Eskolta School Research & Design, New York, NY, USA;* [c]*University of California, Berkeley, CA, USA*

Engaging in science as an argumentative practice can promote students' critical thinking, reflection, and evaluation of evidence. However, many do not approach science in this way. Furthermore, the presumed confrontational nature of argumentation may run against cultural norms particularly during the sensitive time of early adolescence. This paper explores whether middle-school students' ability to engage in critical components of argumentation in science impacts science classroom learning. It also examines whether students' *willingness* to do so attenuates or moderates that benefit. In other words, does one need to be both willing and able to engage critically with the discursive nature of science to receive benefits to learning? This study of middle-school students participating in four months of inquiry science shows a positive impact of argumentative sensemaking ability on learning, as well as instances of a moderating effect of one's willingness to engage in argumentative discourse. Possible mechanisms and the potential impacts to educational practices are discussed.

Keywords: *Science learning; Argumentative sensemaking; Middle school*

1. Introduction

In an effort to curb the trend of declining student performance and motivation in the sciences (Bathgate, Schunn, & Correnti, 2014; Evans & Stanovich, 2013; Glynn, Brickman, Armstrong, & Taasoobshirazi, 2011; Osborne, Simon, & Collins, 2003; Vedder-Weiss & Fortus, 2012), educators and researchers are exploring approaches

*Corresponding author. Learning Research & Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260, USA. Email: meb139@pitt.edu

that promote student engagement in meaningful interactions with science content including experiences that are more closely aligned to the authentic practices of science (Bricker & Bell, 2008; Herrenkohl & Guerra, 1998). Such activities include inquiry-based, problem-focused tasks that encourage group discourse and sharing of ideas (Berland, 2011; Bricker & Bell, 2008; Cohen & Ball, 2001; Osborne, Erduran, & Simon, 2004; Tabak & Baumgartner, 2004) and, as such, resemble the practices of professional science. Recent standards in science education have also reflected such a shift to include the practices of professional science (e.g. the Next Generation Science Standards in the US; Achieve, 2012).

Argumentation (e.g. discussing evidence and debating ideas) is a critical element in authentic science-learning experiences and is vital in professional science training (Bricker & Bell, 2008; Duschl & Grandy, 2008; Kuhn, Kenyon, & Reiser, 2006; Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003). Engaging in argumentation encourages critical thinking fundamental to the practice of science. Discourse encourages students to consider alternative viewpoints and evidence which leads to questions that allow one to gather requisite knowledge to formulate an informed opinion (Brickner & Bell, 2008; Chin & Osborne, 2008). A number of studies, however, have found that many students have only rudimentary argumentation skills (Kuhn et al., 2006; Lemke, 1990; Osborne et al., 2003). Despite having weak argumentation skills, research continues to show that students benefit from using discourse in their education. Students encouraged to discuss and debate, rather than to simply combine answers, show greater learning scores and greater maintenance of this knowledge over time (Asterhan & Schwarz, 2007, 2009). A recent meta-analysis (Chi, 2009) classifying and examining activity structures showed activities with interactive properties (e.g. involving substantive dialogue with peers/partner including responding to feedback, arguing and defending answers, and incorporating multiple points of view) showed the greatest learning gains for students.

Yet the problem in engaging students in such beneficial dialogue is more than lack of argumentation skills; students may also be unwilling to use the skills they do have. Argumentation can be confrontational, which may run against cultural norms or personal preferences, particularly for early adolescents, who are often working hard to fit in with peers to maintain social hierarchies in the school (Aikins, Bierman & Parker, 2005; Anderman, 2003; Buhrmester, 1990; Eckert, 1989). Yet, there is limited research to date related to students' perceptions of the potential social risks of engaging in such discourse (e.g. disagreeing with a peer). Little is known about both whether social risks of discourse are strongly felt in science classroom contexts and whether such perceived risks limit student learning in science classrooms.

Given the beneficial effect argumentative discourse has on learning combined with the limited research on how young students perceive engaging in such practices, this study has two goals. First, we examine whether students benefit from the ability to engage effectively in scientific argumentation. That is, does capability with the processes of professional science impact learning scientific content in science classrooms? Second, we investigate why students might be unwilling to engage in argumentation in science and whether willingness to engage in argumentation in science moderates the benefits of having the ability.

1.1. *Capability: Why is Ability to Argue Critical to Science Learning*

Science as a discipline is often characterized as fundamentally argumentative in nature:

> There is no abstract and logical scientific method apart from the actions of scientists and engineers … scientists and engineers are always in the position of having to convince their peers and others of the value of their favorite ideas and plans—they are constantly engaged in struggles to gain resources and to promote their views'. (Sismondo, 2004)

Many researchers have also claimed that argument is foundational to science learning as well (Berland, 2011; Bricker & Bell, 2008; Cohen & Ball, 2001; Osborne et al., 2004; Tabak & Baumgartner, 2004). Indeed, Means and Voss (1996) position argumentation as the means by which informal science knowledge is formed and has been further established as a key purpose for human reasoning (Chater, Oaksford, Hahn & Heit, 2010; Mercier & Sperber, 2011). But empirical evidence on the benefits of argumentation ability for science learning is still lacking; do students with such abilities actually learn more science content than students who do not have such abilities?

In order to effectively engage in scientific argumentation, two aspects are likely needed, which collectively we call *argumentative sensemaking*: (1) The ability to structure an argument, which includes providing support for one's position and taking on opposing views and (2) The use of the reasoning of science, which is critical for selecting appropriate support, generating relevant opposing views, and selecting appropriate counterarguments to the opposing views. In science, this reasoning involves the generation of relevant evidence, informed interpretation of evidence, reasoning through confounds, as well as understanding phenomena in relation to causal mechanisms. These details of these components of argumentative sensemaking are what distinguish argumentation in science from argumentation in other domains (Berland & McNeill, 2010; McNeill, 2011). We now review skills likely to be critical in scientific argumentation.

Generating investigable questions. Students' generation of investigable questions (i.e. questions that direct learning and are responsive to first-hand inquiry) (Chin & Kayalvizhi, 2002; Chin & Osborne, 2008) can serve many purposes in science classrooms (e.g. promote critical reflection and self-evaluation of knowledge, encourage interest in science topics) (Chin, 2006; Watts, Alsop, Gould, & Walsh, 1997) and is considered one of the central features of scientific inquiry as well as a major component in scientific discourse (Cuccio-Schirripa & Steiner, 2000). Chin and Osborne's (2008) review of the categories and functions of student questions demonstrates the power of accessing these questions as a way to capture depth of student understanding and can be used to teach the nature of scientific investigation and discourse in the professional field.

Evaluating evidence. Evidence can be defined as information used in forming a conclusion or judgment—the data gathered and combined to construct and defend claims/explanations within an argument. One evaluates evidence in light of the question being asked or argument being made. In science, one of the primary sources of evidence is from empirical data. Critical component abilities include: (1) extracting relevant data from investigations appropriate for the research question; (2) analyzing and interpret graphs, tables, and data accurately and with intention; and (3) interpreting the outcomes of controlled and comparative experiments to answer questions (Erduran & Jiménez-Aleixandre, 2008; Sampson & Clark, 2006; Toulmin, 2003).

Understanding mechanism. Mechanistic reasoning is a central focus of explanations/justifications in science. We define sophisticated causal reasoning as that which posits a scientifically accurate mechanistic explanation for how a particular cause brings about an effect (Koslowski, 1996; Schauble, 1996), and these mechanisms may often be domain-specific (Carey, 1995). Literature advocating the use of mechanisms to explain phenomena generally describes them as identifying the process between causes and effects.

Justifying argument. Argumentation is the ability to support claims with evidence and justify the validity of the relationship between claim and evidence within a universe of alternatives. The argument can be advanced in writing, in a speech, or in an unstructured dialog. Argumentation is growing in importance as a tool for science learning in the classroom (Bell, 2004; Osborne et al., 2004), in addition to being a core component of authentic scientific thinking (Toulmin, Rieke, & Janik, 1984) and a critical focus of the large-scale standards reform effort—*Common Core Standards* in both English language arts and mathematics as well as Next Generation Science Standards—currently underway in the USA. Engaging in argumentative activities in classroom science has been shown to develop scientific thinking (Koslowski, 1996; Kuhn, 1992, 1993) and promote conceptual change (Andriessen, 2006).

Also within argumentation, the social context of alternative views is an important aspect of how science knowledge develops. Scientists are always pitting their own ideas against other potential explanations for the same phenomenon, and it is through argumentation that ideas become accepted, rejected, or modified. Many researchers claim that without an understanding of the context of a scientific argument, one cannot fully understand the ideas being presented; that is, no ideas are born or expressed in a vacuum, they are always contextualized within a larger social and cognitive context (Billig, 1987; Collins & Pinch, 1993; Pera, 1994; Toulmin, 2003). Therefore, we include an ability to contextualize an argument in the context of opposing views as an element of argumentative sensemaking.

Unlike prior research that has examined each of these elements in isolation, we hypothesize that the combination of these skills produce an overall argumentative sensemaking ability that can enable students to more effectively acquire scientific knowledge from those experiences. In other words, argumentation around scientific content involves a number of critical sub-skills that work together, and when they work together they enable learning of scientific content from experiences. Our study will test this critical hypothesis.

1.2. *Willingness: Perceptions of the Risks and Values of Discourse*

We argue that ability in argumentative sensemaking is not the only driver as to whether students participate effectively in science-learning experiences. As science is intrinsically argumentative, student perceptions of themselves and their peers as members of a discourse community are likely to influence how students approach science learning. Qualitative work with adolescents who were asked about their perceptions of the purposes (i.e. the benefits) and risks (i.e. the potential negative social perception) of peer-to-peer argumentative discourse revealed that students indicated that such discourse is important because it can lead to learning, persuasion, or self-expression (Kuhn, Wang, & Li, 2010). For example, in Kuhn et al.'s (2010) study, when sixth-grade students were presented a scenario in which they disagreed with their friend about who should be elected governor of their state and asked whether they would discuss the disagreement with their friend, some students saw such discussion as beneficial for understanding (e.g. 'Because she might know more about the other candidate and I might now more about the one I'm for. We could tell each other more and we would learn something new[…]'; 'Because if you discuss the candidates' pros and cons you might see that either you or your friend is for the wrong person and might want to change sides' p. 34).

Yet, students often see risks of discourse, in general, and of argumentation, more specifically. For example, these same students in the Kuhn et al. study worried that disagreeing with a peer could have negative social consequences (e.g. 'You could get into a serious argument and you may not be friends anymore. Also, if you don't [discuss your disagreement], arguments can be avoided and your friendship has a better chance of being longer', p. 34). Adolescence is a time of particular sensitivity to social risks that may threaten one's sense of belonging and relatedness to one's peers (Baumeister & Leary, 1995; Eckert, 1989; Ryan & Deci, 2000; Vallerand, 1997). A large body of work has shown that students' quality of friendship, their sense of relatedness, and their emphasis on peer acceptance are keenly felt by students of this age and are associated with both social (e.g. anxiety, depression, self-esteem) and academic outcomes (Aikins et al., 2005; Bellmore, 2011; Kingery, Erdley, & Marshall, 2011; Waldrip & Jensen-Campbell, 2008; Wentzel, Barry, & Caldwell, 2004). In fact, social concerns can outweigh academic goals for these students (Maute & Brough, 2002). Therefore, they may avoid disagreements due to the potential risk of peer rejection.

In sum, students often recognize benefits of discourse, but there are many relevant risks to discourse as well. How students perceive and balance the purposes and risk of discourse, particularly in the case of disagreement, would likely impact how willing they are to engage in such discourse within the classroom and with peers. Therefore, further research is needed to understand students' sensitivity toward benefits and risks: Are benefits of argumentation in science broadly acknowledged by students? Which risks are mostly broadly felt? Which risks are barriers to science learning? Note that apprehension about peer-to-peer argumentation could influence overall views of argumentation and thereby broadly influence skill and knowledge development in science.

1.3. *The Current Study*

In this paper, we first examine whether middle-school students' argumentative sensemaking is related to their classroom science learning. Specifically, we ask three research questions (RQ):

- RQ1. Does argumentative sensemaking (ability), that is, treating scientific situations as argumentative, empirically grounded, and mechanistic (Driver, Leach, Millar, & Scott, 1996), support learning in a middle-school inquiry science classroom?
- RQ2. Does willingness to engage in argumentative discourse predict middle-school science learning?
- RQ3. Does one's willingness to engage in argumentative discourse moderate the effectiveness of argumentative sensemaking on learning? Specifically, does argumentative sensemaking support learning only for middle-school students who are willing to participate in potentially socially risky, disagreeing classroom discourse?

2. Methods

2.1. *Participants*

Recruited participants were enrolled in 34 sixth-grade science classes from 10 urban public schools in a mid-size city in the Eastern United States. Science teachers from the schools were invited to participate in the study at a school district in-service event. Almost all teachers present at the event accepted the invitation, and those who accepted were compensated for their time according to the number of participating classes. As a result, the participating schools represented a diverse sample of students by socio-economic status, ethnicity, and achievement scores. Overall, students were primarily Caucasian (33%) or African-American (35%), and 43% came from families in which at least one parent (49% female) had a college education. But particular schools ranged from being majority Caucasian to almost exclusively African-American, and had similar diversity in percentage of students eligible for free or subsidized lunch.

Sample size varied across measures largely due to student absence across multiple data collection points: 858 students completed the Argumentative Sensemaking measure, 816 students completed both the pre-test and learning outcome, 607 completed the purposes and risks of discourse items, and 441 completed demographic information.[1] More specific sample size information is included in each analysis. Incentives for students included the opportunity to win $200 (one winner per class), and a self-chosen activity kit (e.g. mosaic art kit, simple machine kit) for each consenting participant.

2.2. *Assessments*

2.2.1 *Argumentative sensemaking.* We developed a measure of students' ability to make effective arguments in science. As argumentation in science takes place within a content domain, the measure was situated in a particular content scenario. However, the measure made use of science content that was accessible to a range of students with varying prior knowledge in order to measure ability to engage in various aspects of scientific reasoning and argumentation, rather than topic interest or prior experience with particular learning domains. Specifically, the instrument focuses on a topic of broad interest to middle-school students (Bathgate et al., 2014)—how best to assist a population of pink dolphins who are threatened with extinction. There was enough information embedded into the scenario to ensure that even a student who was relatively unfamiliar with the topic would have the opportunity to perform well.

This measure, developed in concert with a biologist who studies dolphin habitats, was piloted and revised across multiple rounds with 11 fifth- and sixth-grade students through cognitive interviews. These cognitive interviews involved students meeting 1:1 with a trained researcher and reading each item aloud, restating it in their own words, and articulating the reasons for choosing their response.

The resulting nine-item instrument captures the aspects of argumentative sense-making through seven multiple-choice items and two open-ended items. The scenario culmination (and final response item) involves having the student write a letter to President Obama, who (the students were told) was set to attend a meeting focused on whether the dolphins' natural habitat ought to be improved or whether the dolphins should be moved to a protected cove (Table 1). Prior to starting the letter, students were asked to select whether they felt the dolphins should be moved to the cove or whether the dolphins' natural habitat could be improved (evidence presented throughout the scenario did not generally favor one choice over the other, to allow for legitimate argumentation and factors to consider on each side of an argument).

The written letter that students produced in response to the measure prompt was coded with two coding guides: one capturing how well the student justified his or her claims with evidence and reasoning, and one that captured how well the argument was contextualized in the provided alternatives (i.e. how well they understood and refuted the opposing position). Thus, each student had two coded variables for this item (see Tables 2 and 3 for coding scheme and examples). The second open-ended question was coded as to how well students developed investigable questions (Table 4). The overall instrument had acceptable reliability (alpha = .77).

2.2.2. *Willingness measure: purposes and risks of argumentative discourse.* Students' perception of the purposes and risks of argumentative discourse in science were assessed with six questions contextualized in the same scenario as the scientific sense-making assessment. Students were asked 'You and your friend disagree about what should be done to help the dolphins. Would you talk to your friend about your ideas?' Six responses drawn from work on young adolescents' perceptions of the purposes and risks of argumentative discourse were provided (Kuhn et al., 2010) and students were asked to indicate the extent to which they agreed with each (choosing from

Table 1. Sample questions from the argumentative sensemaking measure (representing the 'protected cove' option) organized by dimension

Dimension	Desired responses	Sample questions
Questions	Pose investigable questions and/or identify investigable problems	If you could ask a dolphin expert any questions to support the idea that it is better to bring the dolphins into the cove to protect them rather than improving their natural habitat, what would you ask? List at least three detailed questions *Response coded for investigability of generated questions (Cohen's Kappa = .87)*
Evaluation and Utilization of Evidence	Evaluate relative quality of evidence	Maria is wondering which dolphin has the most spots. What is the best evidence she could get to answer her question? (a) She could guess which dolphin has the most spots (b) She could choose a dolphin and count the number of dark spots and compare to the number of light spots (c) She could ask her friends which dolphin looks like it has the most spots (d) *She could count the number of spots on all of the dolphins*
	Utilize provided evidence in support of written argument	Now, write a letter to President Obama explaining why you think that moving the dolphins to a protected cove is the best way to help them. As you write your letter you can use the provided evidence and any other information that you know to support your argument. *Response coded for justification of ideas utilizing reasoning and evidence (Cohen's Kappa = .88)*
Explanations	Construct mechanistic explanations of phenomena	You said that X makes dolphins leave the coves. How could X make dolphins leave the cove? Please explain how X can affect dolphins *Response coded for mechanistic reasoning (Cohen's Kappa = .88.)*
	Contextualize response in provided alternatives	Now, write a letter to President Obama explaining why you think that moving the dolphins to a protected cove is the best way to help them. As you write your letter you can use the provided evidence and any other information that you know to support your argument *Response coded for contextualization of response in alternative (Cohen's Kappa = .93)*

Table 2. Justification coding rubric with example responses

Coding criteria	Student responses
(5): Claim connected to reasoning and evidence	'Dear President Obama, My name is [deleted] and I think that you should try to improve pink dolphin's natural habitat instead of moving them to a different spot. Here are some reasons why. Pink dolphins have been in their natural habitat for a long time. They have adjusted to that environment and probably use it for living. They could use their environment for finding food. Pink dolphins probably have adjusted to that spot because it has the kind of food that they eat. They probably need that environment to survive too. They could've adjusted to that weather and have been around more than boats and trash has been there. I think that if you make that area a natural habitat for pink dolphins that no fishing boats or trash or lots of noise can enter in it then that would help increase the population of the dolphins. But if you take them to a different cove they might go away in change of weather and probably because they don't know where they are. Best Regards'
(4): Both reasoning and evidence	'It hurts the animal to move them away from their habitat if you take an animal out the animal that eats it wont have food and the population their will die and the amount of the animal it eats will over populates. That would make the animals fell much safer'
(3): Unsubstantiated reasoning OR Unwarranted evidence	'When the dolphins are being moved they cannot breath they will die, so can you just clean up for them? And make their water better?' 'I think natural habitat is my Best way to help them because they might have a hard time to adjusting to their new environment'
(2): Solutions only	'We should start to clean streams, rivers, lakes and oceans' 'Dear Obama, You should tell people to stop throwing trash to the dolphin natural habitat and people marking noisy sound and people on boat making loud, noisy sound and save the dolphin natural habitat'
(1): Personal OR Severity only	'IT'S THE BEST TO HELP THEM BECUASE DOLPHNS ARE AWESOME' 'Find out a new way to find to save our dolphins, because one day we'll need them. I really don't know what else to say'
(0): Claim only	'I think you should move them instead of keeping them there'

Note: Minor spelling and grammar errors of students have been corrected for clarity.

completely agree, agree, disagree, and completely disagree). Three of the statements represented the purposes of argumentative discourse:

- Yes, because maybe you could learn something from your friend that you did not know before.
- Yes, because maybe you could convince your friend to see things your way.
- Yes, because it is important to share your opinions.

Table 3. Duality coding rubric with example responses

Coding criteria	Student responses
(4): Integrated Perspective (+Other OR −Own)	'I think that people should improve their habitat because it will be very hard for them to adjust to the new environment. I know it is for a good cause, but they don't know where they are taking them, so this will make them really scared. Would you be scared if someone took you but you didn't know where you were going? I don't think they should be put through this hard transition' 'Dear Barack Obama, I think that moving the dolphins to a protected cove is the best way to help them. By moving them it won't make a threat to extinction But it will take some time for them to adjust to their new environment. They will not face with all the trash, noise, and the boats'
(3): Dual Perspective (−Other and +Own)	'Dear president Obama, I think you should just make it where no trash, boats, and noise into their habitat. If you take them out they won't be able to know where to hunt or if the area is safe and they will move probably into shark territory'
(2): One-sided argument (−Other)	'moving the dolphins could just make it worse because there might be more noise there then in their old habitat'
(1): One-sided argument (+Own)	'It is the best way to protect them because their would be less things there. And because so they can be more safe. Also so there would be less noise'
(0): No strengths or weaknesses in argument	'Dear President Obama, I strongly intend and will not give up until I get an answer. We need to help the pink dolphins. And give them a clean place to leave. -Your friend'

Note: Minor spelling and grammar errors of students have been corrected for clarity.

Table 4. Coding criteria for type of questions

Coding criteria	Student responses
(3): Asks informative and investigable questions	'Would the dolphins be more confortable in captivity or their natural habitat? Does the decreasing of the noise, boat traffic, and trash have an effect on the dolphins in their natural habitat? Would the captivity not only make them safe from the danger of the wild, but also make them more comfortable than in their natural habitats?'
(2): General questions	'Can you give the dolphins cleaner water? Can you clean up the trash? Can you stop the boats from going over the dolphins?'
(1): Irrelevant questions OR Statements with relevant information	'Do dolphins have teeth? How munch do dolphins swim? Do dolphins sleep?' 'We should tell people to stop dumping stuff in the water'
(0): No response or off topic	'I don't know because I don't like dolphins'

Note: Minor spelling and grammar errors of students have been corrected for clarity.

Three of the statements represented the risks associated with argumentative discourse:

- No, because you cannot change another person's opinion.
- No, because it could hurt your friendship.
- No, because it is not nice to disagree with people.
- A description of how these items were coded and then scored is presented in the results section of this paper.

2.2.3. *Science learning*. Learning of academic content in the classroom was assessed using a knowledge test administered before and after the four-month classroom unit on weather and climate. This unit consisted of a series of 'investigations' to examine different factors related to weather and water (e.g. 'Air Pressure and Wind'). The goals of the unit and its sections were not related to argumentation but rather focused on developing content knowledge (e.g. Objective: 'Pressure exerted on a gas reduces its volume and increases its density') through inquiry (e.g. Objective: 'Apply pressure to a system and observe the compression of the gas') and explanation (e.g. Objective: 'Describe the relationship between changing air pressure and wind'). Activities included daily warm-up writing exercises, often a hands-on experiment completed in groups, and some group discussion. Although these discussions afforded opportunity for discourse, there was no direct preparation or assessment of discourse provided by the curriculum. Prior work has emphasized the difference between discussion that is argumentative vs. collaborative in nature (Asterhan & Schwarz, 2007, 2009) and the Full Option Science System™ (FOSS) curricula is focused on the latter, as is commonly done in science curricula. Classroom discussions generally answered a structured series of questions (e.g. 'Is the volume inside the plastic jar greater, the same, or less when you squeeze it?') and group discussions focused on explaining an observed result. Thus, there are opportunities for students to include argumentative discourse in these discussions, but the students, not the curriculum, would drive the nature of the discourse.

Many students have some prior knowledge about weather and climate, and this prior knowledge could be correlated with argumentation ability and/or attitudes toward argumentative discourse because of support at home or other informal learning experiences on science practices and science content. Therefore, in order to assess the amount of learning that occurred during the study duration, it is critical to look at growth in knowledge, rather than just final knowledge levels.

The test was built from released Trends in International Mathematics and Science Study, National Assessment of Educational Progress, and state science test items that matched the overall content of the classroom unit (Kuhn & Crowell, 2011). Only those items that matched the parts of the unit covered by all teachers were included in the final analysis (21 items, alpha = .78). Two example items include: (1) 'What is the primary energy source that drives all weather events, including precipitation, hurricanes, and tornadoes? a) The Sun; b) The Moon; c) Earth's gravity; d) Earth's rotation' and (2) 'Locations in the Southern Hemisphere typically experience

___________. a) Longer days in June than January; b) Equal amounts of daylight and darkness; c) Longer days in January than June'. Each item is given a one for the correct choice and a zero for all other choices.

2.2.4. *Additional covariates.* To establish the contribution of our measures above that of general academic ability, students' fifth-grade math ($N = 645$) and reading ($N = 628$) state test scores were obtained. These tests were administered at the end of fifth grade. They have reliabilities above .9, as presented in published technical reports (e.g. Data Recognition Corporation, 2012). Free and subsidized lunch status was also included to control for the broad set of performance factors associated with socio-economic status.

2.3. *Procedures and Context*

In the first week of school, students were given the learning outcome pre-test measure and the argumentative sensemaking measure. After completing the argumentative sensemaking assessment, students were assessed on their perceptions of the purposes and risks of argumentative discourse. On a separate day, demographic information was obtained including age, race, and gender. At the end of the semester, the learning outcome measure was administered again to measure content-learning gains.

All students experienced a common curriculum over this time period: adaptations by the National Research and Development Center on Cognition and Science Instruction (Kuhn & Crowell, 2011) to one earth sciences unit (on the topics of weather and climate) of the widely used FOSS curriculum; the adaptations included specific daily calendar of specific activities for each teacher and thus provided a relatively homogeneous learning environment across teachers and classrooms. Neither the base FOSS curriculum nor these adaptations specifically teaches argumentative discourse, and thus, without intense intervention the skills students had at the beginning of the unit were likely representative of skills they had throughout the unit (Kuhn & Crowell, 2011). Yet, as a hands-on inquiry science curriculum, there were frequent opportunities for students to engage in discourse with peers in small group activities and whole class discussions.

To insure relatively high homogeneity of implementation fidelity, teachers were provided a one-day professional development experience in which the curriculum adaptations were described. Classroom observations and self-report surveys by the teachers suggest that the key elements of the curriculum were implemented as designed, with only some variation in the rate at which the materials were covered.

3. Results

3.1. *Does Argumentative Sensemaking Ability Support Learning in a Middle-school Inquiry Classroom (RQ1)?*

Each student's scientific sensemaking score was computed using item-response theory (IRT). IRT affords the most robust approach for the current purpose, addressing

Table 5. Regression results for content knowledge post-test using pre-test scores, state testing scores, free/subsidized lunch, and argumentative sensemaking scores

Variable	Model 1		Model 2	
	B	*p*	*β*	*p*
Pre-test score	.32	<.001	.28	<.001
Fifth-grade math score	.15	.003	.13	.013
Fifth-grade reading score	.19	<.001	.15	.004
Free/subsidized lunch status	.01	.716	.02	.611
Argumentative sensemaking score			.15	.001

Note: All predictors were standardized using a *z*-score prior to analyses.

missing items as well as the influence of differential item difficulty on computing a final score (Sondergeld & Johnson, 2014). To test whether argumentative sensemaking predicts science learning, we ran linear regressions predicting the knowledge post-test. In model 1, the model only included pre-test, fifth-grade state standardized testing scores for math and reading, and students' free and subsidized lunch status. Argumentative discourse scores were added in model 2. As Table 5 shows, students' argumentative sensemaking ability predicts science content learning beyond standardized test scores. Note that the same results are obtained with more complex hierarchical models taking into account the nesting of students within classrooms and teachers.

Similar results are found using simpler statistical models: in a two-variable linear regression, both the content pre-test ($F(1, 722) = 94.4$, $p < .001$) and argumentative sensemaking scores ($F(1, 722) = 63.0$, $p < .001$) were strong predictors of post-test. For simplicity of reporting, this model is used as the foundation for the analyses in the next section, although similar results are also obtained from more complex models.

3.2. *Does Willingness (Perceptions of Discourse) Predict Middle-school Science Learning (RQ2)?*

Having established scientific sensemaking's relationship to learning, we now explore the additional influence of students' perception of discourse items. Table 6 shows the average responses for each of the six items on the 1–4 scale, along with distributions across a binary split of agree vs. disagree (indicted by % endorsed). As can be seen by the means and distribution, students overwhelmingly endorsed the purposes of discourse. As found by Kuhn et al. (2010), most middle-school students are already aware of the personal benefits of discourse in science. The risks were endorsed less frequently than the benefits, and there was considerable variation on the extent to which each risk was endorsed. However, a majority of students endorsed at least one risk and even the least felt risk was endorsed by more than a quarter of the students.

Because almost all students endorsed the benefits of discourse items, there is no sufficient power to analyze effects of perceived discourse benefits on learning. By contrast,

Table 6. Mean and standard deviation of ratings and % of students endorsing each potential benefit and risk of discourse in science

Category	Item	*M*	SD	% endorsed
Purpose	Learn something new	3.4	0.8	89
Purpose	Convince friend	3.2	0.8	85
Purpose	Important to share	3.4	0.8	89
Risk	Cannot change opinion	2.7	1.1	58
Risk	Hurt friendship	2.4	1.0	44
Risk	Not nice to disagree	2.0	1.0	27

there was sufficient student variation in their perceptions of the risks associated with discourse and their association with content learning was further analyzed.

First, we present the learning gains associated with endorsing or rejecting each risk of discourse item. As seen in Figure 1, rejecting the risks (in other words, not feeling there is a risk) associated with each item is related to higher science post-test scores as opposed to students who endorsed the risk using simple *t*-tests.

However, this analysis does not address the relationship among the items and whether each risk contributes independently to learning gains. In other words, is each risk independently predictive? Further, are they predictive controlling for argumentative sensemaking and prior knowledge? A multiple regression was run on the science post-test scores including each risk of discourse item (see Table 6 for means and standard deviations), argumentative sensemaking score ($M = 0.04$, $SD = 0.65$), and pretotal scores ($M = 14.0$, $SD = 4.8$). None of the variables are so strongly correlated that collinearity problems are an issue (Table 7).

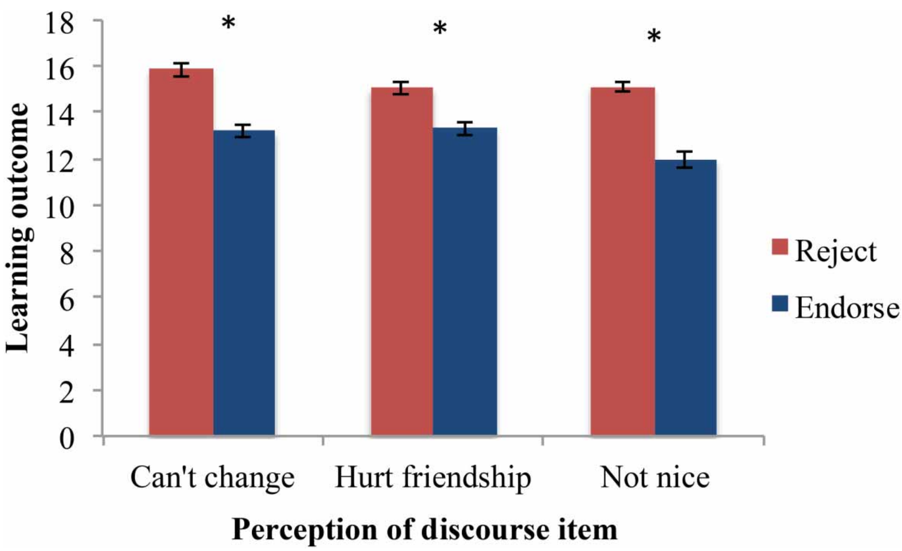

Figure 1. Mean knowledge post-test scores (and SE bars) among students who endorse or reject each of the risk items associated with discourse (*$p < .05$)

Table 7. Correlations among risk ratings, argumentative sensemaking, and knowledge pre- and post-test scores

	Hurt friendship	Not nice to disagree	Argumentative sensemaking	Pre-test	Post-test
Cannot change opinion	.32	.29	−.26	−.26	−.26
Hurt friendship		.43	−.25	−.19	−.18
Not nice to disagree			−.32	−.27	−.26
Argumentative Sensemaking				.45	.44
Pre-test					.45

Note: All correlations are significant at the $p < .001$ level. Sample size varies depending on comparison, ranging from 561 for post-test comparisons to 777 for argumentative sensemaking and pretotal scores.

Regression results (Table 8) show that in addition to pre-test and argumentative sensemaking, two of the risk perceptions associated with discourse are independently predictive of learning outcomes. Endorsing the risk that engaging in discourse is useless as you cannot change another's opinion and/or endorsing the risk that disagreement is 'not nice' leads to significant lower classroom learning compared to rejecting these risks. Additionally, the low correlations of these risks with each other and with argumentative sensemaking ability confirm the independence of these items as predictors.

3.3. *Does Ability in Argumentative Sensemaking Predict Learning only if Students are Willing to Reject Risks Associated with Discourse (RQ3)?*

To explore whether perceptions of risks associated with discourse moderates argumentative sensemaking ability on learning, separate ANCOVAs with knowledge post-test as the dependent variable were run for each risk item, controlling for argumentative sensemaking scores and knowledge pre-test scores. An interaction term examining risk X argumentative sensemaking was included. In other words, each model examined whether there was an interaction between students who endorsed

Table 8. Multiple regression results for knowledge post-test scores predicted by risk perceptions, scientific argumentation, and pretotal scores

Variable	B	p
Pre-test	.31	<.001
Argumentative sensemaking	.28	<.001
Cannot change opinion	−.09	.021
Hurt friendship	.01	.251
Not nice to disagree	−.12	.002

Note: Sample size for regression=571.

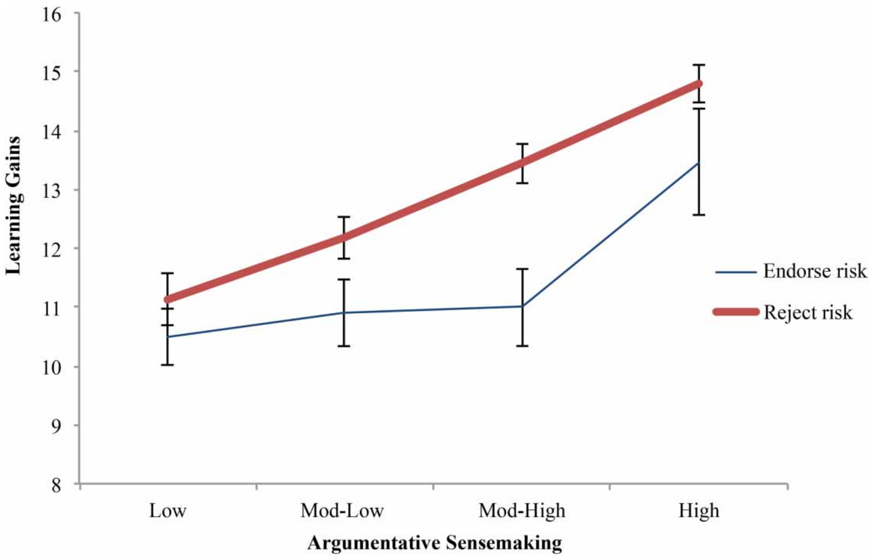

Figure 2. Mean knowledge post-test scores (and SE bars) as a function of endorsing the 'it is not nice to argue' risk of discourse and argumentative sensemaking scores

the item (e.g. agreed) or not (e.g. disagreed) and argumentative sensemaking scores in predicting knowledge post-test scores (while controlling for pre-test).

For all three items models, pre-test and argumentative sensemaking remained strongly predictive of learning ($p < .001$ in all models). Similar to the results of the prior multi-regression analyses, there was a main effect for the risk item 'you can't change another person's opinion' ($F(1, 531) = 8.2$, $p = .004$) and 'it's not nice to disagree with people', ($F(1, 531) = 14.9$, $p < .001$), but not for the 'you might hurt your friendship' item ($F(1, 532) = 1.2$, $p = .27$).

Only the risk that 'it's not nice to disagree with people' had a statistically significant interaction effect ($F(1, 531) = 3.8$, $p = .05$). Figure 2 shows the interaction for the 'not nice' risk item across four levels of argumentative sensemaking (binned into quarters: low, moderately low, moderately high, and high argumentative sensemaking). This figure shows that students' perception of the social risk (being perceived as not nice if one disagrees) impacted learning gains for all except the lowest performing argumentative sensemaking group. Alternatively, it also shows that increasing argumentative sensemaking scores had little benefit for student learning when they endorse this particular risk of discourse in science.

4. Discussion

4.1. *Willing & Able: Argumentative Sensemaking and Willingness to Engage in Discourse Affect Learning Independently and in Combination*

Our current study revealed a strong, positive relationship between argumentative sensemaking and learning. First, students who had more of the central abilities involved in argumentation in science (e.g. effective evidence evaluation, mechanistic reasoning, justification, duality of argument) were able to learn more science knowledge content from their class than their peers with fewer such abilities. Such findings add weight to calls for truly integrative science instruction in which students develop skills in the practices of science together with learning the body of knowledge of science; here we show that such skills critically support content learning rather than simply embodying the justifications of science knowledge.

Furthermore, we show that how one perceives the risks associated with argumentative discourse influences science learning and, in the case of perceiving disagreement as 'not nice', also strongly moderates the benefit of argumentative sensemaking ability on learning. When students prefer not to discuss their ideas because it is not nice to disagree, then students appear to benefit little (at least in terms of content learning) from having higher levels of sensemaking skills. By contrast, those who reject this risk see significant gains in classroom learning across the spectrum of sensemaking skill levels, with only the lowest sensemaking skill levels showing similar gains as those who endorse the risks of argumentative discourse.

When interpreting these results, it is important to consider multiple causal possibilities. We begin with the modest but significant relationship between being willing and being able to engage in scientific argumentation. It is possible that being more comfortable with disagreement supports the development of argumentative sensemaking skills; alternatively, it is possible that having a greater propensity toward argumentative sensemaking paves the way for comfort with disagreement. We argue that these relationships are most likely reciprocal, with greater sensemaking helping one to be comfortable with disagreement and further comfort allowing for sensemaking to develop. However, detailed longitudinal research would be required to test this hypothesis directly, particularly within diverse cultural contexts.

Next, there is the relationship between being able to engage in scientific argumentation and content learning. It is possible that content knowledge plays a role in supporting argumentation in addition to the currently made claim that ability to engage in scientific argumentations supports content learning. In other words, students with greater content knowledge may have a more developed view on a given topic or perceive themselves more capable of engaging in effective argumentation, which may in turn affect their willingness to take part in discussion. We used an approach to test ability using contents that are generally available to students to minimize the possibility of a reversed causality. In addition, the statistical analyses controlled for prior content knowledge in the domain. Nonetheless, some reverse causality is possible and many prior researchers have argued that rich domains in which students have knowledge of the content are important for building argumentation skills (e.g. Kuhn & Crowell, 2011).

Why would argumentation (willingness and ability) play such a large moderating role in classroom content learning in science? Argumentative discourse in the science classroom can derive power from at least three sources: the fundamentally argumentative nature of science, the cognitive complexity of argumentation, and the particularly social nature of adolescence. First, in the relationship between argumentation and learning generally, there is a particularly strong case for how argumentation supports science learning (Berland, 2011; Berland & Reiser, 2011; Ford, 2012; Koslowski, 1996). Science is a fundamentally discursive endeavor in which the community of scientists construct, support, and evaluate claims (Bricker & Bell, 2008; Norris & Phillips, 2008; Phillips & Norris, 2009; Sismondo, 2004); in short, scientists utilize argumentative practices to negotiate the prevailing explanations for the phenomena that they study. This alignment indicates that there might be a particularly large effect of one's understanding of one's classroom culture on one's science learning.

Next, disagreeing provides the opportunity to explain one's understanding and explaining one's own thinking can produce large learning benefits (Chi & VanLehn, 1991; McNamara, 2004). In particular, argumentative discourse provides the opportunity to reflect upon and elaborate one's ideas (Andriessen, 2006), as well as evaluate one's initial misconceptions (Chinn & Brewer, 1993), which support concept learning and conceptual change (Asterhan & Schwarz, 2009; Bransford, Brown, & Cocking, 1999). As such, the learning benefits related to one's comfort with argumentative discourse presented here could arise from such opportunities for reflection and evaluation through disagreement.

Finally, the fundamentally discursive nature of science is of considerable importance during middle school. Adolescence is a period marked by concern for what one's peers are doing and how those peers perceive one's self (Buchmann & Dalton, 2002; Eckert, 1989; Ryan & Deci, 2000; Véronneau & Dishion, 2010). Furthermore, there is evidence of a relationship between peer influence and students' achievement through middle school (Véronneau & Dishion, 2010) and beyond (Dishion, Nelson, & Bullock, 2004). The social sensitivity of adolescence could explain why the socially oriented items regarding the risks of discourse were powerfully related to scientific sensemaking and learning. These results provide compelling evidence for the power of discourse in classrooms, regardless of whether the curriculum is designed to elicit such discourse.

4.2. *Implications for Education*

Given that perception of argumentative discourse impacts the extent to which argumentative sensemaking facilitates science learning, teachers and practitioners would do well to encourage approaching science as an argumentative endeavor, specifically addressing the role of discourse in science. While our findings show middle-school students are familiar with and endorse the benefits associated with argumentative discourse, teachers should take care to minimize the perceived risks associated with argumentative discourse as these perceptions provide the highest barrier to learning. For example, a classroom environment could encourage the concept that scientific discourse involves disagreement with ideas rather than conflicts among individuals (e.g. scientific *argumentation* is different than *arguing* about personal preferences or beliefs with your best friend, your sibling, your mother, etc.). Even if teachers do not utilize argumentative or discursive activities in their classroom, they should work to ensure that disagreeing discourse is not in opposition with the classroom culture. Encouraging discussion with peers while minimizing the risks associated with such discussion may be particularly beneficial due to adolescents' heightened social sensitivity (Aikins et al., 2005) and the importance of maintaining science interest and achievement during adolescence (Tyson, 2011). Technology-mediated discussions (e.g. via web-based anonymous peer review) might provide some support for learning to argue about ideas rather than with personal stances.

It is important to note that these effects were seen in a science classroom utilizing a scripted inquiry curriculum without a specific focus on argumentation. The relationship between argumentative sensemaking, perceptions of discourse, and learning gains may differ when science learning occurs in these other curricular contexts or with other ages. For example, a classroom focused entirely on textbook learning with no discussion might see no benefits for willingness or ability to engage in scientific argumentation. Alternatively, a classroom with rich supports for argumentation may overcome lack of ability or initial willingness among students.

4.3. *Future Directions for Research*

There remains an open question of how students' perceptions of discourse directly impact their classroom learning in a way associated with sensemaking. What are the mechanisms through which these perceptions influence classroom behavior directly? There are many possibilities: Perhaps students who perceive risks of argumentative discourse experience anxiety or threat reactions when argumentative discourse occurs, and these anxiety reactions reduce learning. Alternatively, perhaps students who value the purposes and reject the social risks of argumentative discourse are more likely to take risks, such as talking in class more frequently. Perhaps these students perceive their science classrooms as more supportive of sharing ideas and discussing disagreements. Perhaps students who reject risks associated with discourse pursue conflicting ideas differently (i.e. students rejecting risks are comfortable with reconciling opposing views or contrasting evidence in a way students sensitive to the risks are not). Additionally, perhaps the consequences of rejecting these risks involve different internal processing rather than only behavioral changes within a class (e.g. a child pursues his or her ideas until satisfied with their conclusion rather than concluding they simply do not know the answer).

Investigations into the mechanisms underlying these differences should be conducted to explore how students' perceptions of argumentative discourse are enacted within a classroom setting. Specifically, it may be necessary to assess particular dimensions of the classroom climate (e.g. frequency and type of peer-to-peer discussion) in order to understand how norms around argumentative discourse are communicated from teacher to students and among students themselves. It likely will also be useful to use interview methods to provide greater detail on the nature of student perceptions of the risks. Design-based research focused on learning more about the features of science-learning experiences that minimize perception of the risks of argumentation would also contribute to building critical knowledge in this area. Detailed examinations of students' perceptions of argumentative discourse, argumentative sensemaking, and classroom interactions may help establish a more direct and causal picture of argumentative discourse in the middle-school science classroom.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funding for this work was provided by The Gordon and Betty Moore Foundation.

Note

1. The reduced sample size for demographic items results from their placement at the end of the assessment to reduce stereotype threat effects [e.g., Steele, 1997]. Due to class time constraints, these items had lower completion rates.

References

Achieve. (2012). *Next generation science standards*. Retrieved from http://www.nextgenscience.org/

Aikins, J. W., Bierman, K. L., & Parker, J. G. (2005). Navigating the transition to junior high school: The influence of pre-transition friendship and self-system characteristics. *Social Development*, *14* (1), 42–60.

Anderman, L. H. (2003). Academic and social perceptions as predictors of change in middle school students' sense of school belonging. *Journal of Experimental Education*, *72*(1), 5–22.

Andriessen, J. (2006). Arguing to learn. In K. Sawyer (Ed.), *Handbook of the learning sciences* (pp. 443–459). Cambridge: Cambridge University Press.

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology*, *99*(3), 626–639.

Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indicators from protocol analyses of peer-to-peer dialog. *Cognitive Science*, *33*(3), 374–400.

Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's motivation towards science across contexts, manner of interaction, and topic. *Science Education*, *98*(2), 189–215.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.

Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 115–143). Mahwah, NJ: Lawrence Erlbaum Associates.

Bellmore, A. (2011). Peer rejection and unpopularity: Associations with GPAs across the transition to middle school. *Journal of Educational Psychology*, *103*(2), 282–295.

Berland, L. K. (2011). Explaining variation in how classroom communities adapt the practice of scientific argumentation. *Journal of the Learning Sciences*, *20*(4), 625–664.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191–216.

Billig, M. (1987). *Arguing and thinking: A rhetorical approach to psychology*. Cambridge: University Press.
Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practices of science education. *Science Education*, *92*(3), 473–498.
Buchmann, C., & Dalton, B. (2002). Interpersonal influences and educational aspirations in 12 countries: The importance of institutional context. *Sociology of Education*, *75*(2), 99–122.
Buhrmester, D. (1990). Intimacy of friendship, interpersonal competence, and adjustment during preadolescence and adolescence. *Child Development*, *61*(4), 1101–1111.
Carey, S. (1995). On the origins of causal understanding. In D. Sperber, D. Premack, & A. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 268–308). Oxford: Clarendon Press.
Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 811–823.
Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105.
Chi, M. T. H., & VanLehn, K. (1991). The content of physics self-explanations. *Journal of Learning Sciences*, *1*(1), 69–105.
Chin, C. (2006). Using self-questioning to promote pupils' process skills thinking. *School Science Review*, *87*(321), 113–122.
Chin, C., & Kayalvizhi, G. (2002). Posing problems for open investigations: What questions do pupils ask? *Research in Science & Technological Education*, *20*(2), 269–287.
Chin, C., & Osborne, J. (2008). Students' questions: A potential resource for teaching and learning. *Students in Science Education*, *44*(1), 1–39.
Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1–49.
Cohen, D. K., & Ball, D. L. (2001). Making change: Instruction and its improvement. *The Phi Delta Kappan*, *83*(1), 73–77.
Collins, H., & Pinch, T. (1993). *The Golem. What everybody should know about science*. Cambridge: Cambridge University Press.
Cuccio-Schirripa, S., & Steiner, H. E. (2000). Enhancement and analysis of science question level for middle school students. *Journal of Research in Science Teaching*, *37*, 210–224.
Dishion, T., Nelson, S., & Bullock, B. (2004). Premature adolescent autonomy: Parent disengagement and deviant peer process in the amplification of problem behaviour. *Journal of Adolescence*, *27*, 515–530.
Data Recognition Corporation. (2012). *Technical report for the 2012 Pennsylvania system of school assessment*. Retrieved from https://www.portal.state.pa.us/portal/server.pt/document/1385966/2013_pssa_technical_report_pdf
Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.
Duschl, R. A., & Grandy, R. E. (2008). Reconsidering the character and role of inquiry in school science: Framing the debates. In R. A. Duschl & R. E. Grandy (Eds.), *Teaching scientific inquiry: Recommendations for research and implementation* (pp. 1–37). Rotterdam: Sense Publishers.
Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. New York, NY: Teachers College Press.
Erduran, S., & Jiménez-Aleixandre, M. P. (2008). *Argumentation in science education*. Florida State University: Spinger.
Evans, J. St. B. T., & Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241.

Ford, M. J. (2012). A dialogic account of sense-making in scientific argumentation and reasoning. *Cognition and Instruction*, *30*(3), 207–245.

Glynn, S. M., Brickman, P., Armstrong, N., & Taasoobshirazi, G. (2011). Science motivation questionnaire II: Validation with science majors and nonscience majors. *Journal of Research in Science Teaching*, *48*(10), 1159–1176.

Herrenkohl, L. R., & Guerra, M. R. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and Instruction*, *16*(4), 431–473.

Kingery, J. N., Erdley, C. A., & Marshall, K. C. (2011). Peer acceptance and friendship as predictors of early adolescence' adjustment across the middle school transition. *Merril-Palmer Quarterly*, *57*(3), 215–243.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, *62*(2), 155–178.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337.

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, *22*(4), 545–552.

Kuhn, D., Wang, Y., & Li, H. (2010). Why argue? Developing understanding of the purposes and values of argumentive discourse. *Discourse Processes*, *48*(1), 26–49.

Kuhn, L., Kenyon, L., & Reiser, B. J. (2006). Fostering scientific argumentation by creating a need for students to attend to each other's claims and evidence. In S. A. Barab, K. E. Hay, & D. Hickey (Eds.), *Proceedings of the seventh international conference of the learning sciences (ICLS)* (pp. 370–375). Mahwah, NJ: LEA.

Lemke, J. L. (1990). *Talking science: Language, learning, and values* (Vol. 1). New York, NY: Ablex Pub.

Maute, J. & Brough, J. A. (2002). The next best step: Helping our students transition 'out'. *Middle Ground*, *6*(1), 16–19.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1–30.

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching*, *48*(7), 793–823.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction*, *14*, 139–178.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.

Norris, S. P., & Phillips, L. M. (2008). Reading as inquiry. In R. A. Duschl & R. E. Grandy (Eds.), *Teaching scientific inquiry: Recommendations for research and implementation* (pp. 233–262). Rotterdam: Sense.

Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What 'ideas-about-science' should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, *40*(7), 692–720.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Pera, M. (1994). *The discourses of science* (pp. 132–33). Chicago, IL: University of Chicago Press.

Phillips, L. M., & Norris, S. P. (2009). Bridging the gap between the language of science and the language of school science through the use of adapted primary literature. *Research in Science Education*, *39*(3), 313–319.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.

Sampson, V. D., & Clark, D. B. (2006, June). Assessment of argument in science education: A critical review of the literature. In *Proceedings of the 7th international conference on learning sciences* (pp. 655–661). Bloomington, IN: International Society of the Learning Sciences.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*(1), 102–119.

Sismondo, S. (2004). *An introduction to science and technology studies*. Malden, MA: Blackwell Publishing.

Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, *98*, 581–613.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629.

Tabak, I., & Baumgartner, E. (2004). The teacher as partner: Exploring participant structures, symmetry, and identity work in scaffolding. *Cognition and Instruction*, *22*(4), 393–429.

Toulmin, S. (2003). *The uses of argument*. New York, NY: Cambridge University Press.

Toulmin, S. E., Rieke, R. D., & Janik, A. (1984). *An introduction to reasoning* (pp. 272–273). New York, NY: Macmillan.

Tyson, W. (2011). Modeling engineering degree attainment using high school and college physics and calculus coursetaking and achievement. *Journal of Engineering Education*, *100*(4), 760–777.

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in Experimental Social Psychology*, *29*, 271–360.

Vedder-Weiss, D., & Fortus, D. (2012). Adolescents' declining motivation to learn science: A follow-up study. *Journal of Research in Science Teaching*, *49*(9), 1057–1095.

Véronneau, M. H., & Dishion, T. J. (2010). Predicting change in early adolescent problem behavior in the middle school years: A mesosystemic perspective on parenting and peer experiences. *Journal of Abnormal Child Psychology*, *38*(8), 1125–1137.

Waldrip, A. M. & Jensen-Campbell, L. A. (2008). With a little help from your friends: The importance of high-quality friendships on early adolescent adjustment. *Social Development*, *17*(4), 832–852.

Watts, M., Alsop, S., Gould, G., & Walsh, A. (1997). Prompting teachers' constructive reflection: Pupils' questions as critical incidents. *International Journal of Science Education*, *19*, 1025–1037.

Wentzel, K. R., Barry, C. M., & Caldwell, K. A. (2004). Friendship in middle school: Influences on motivation and school adjustment. *Journal of Educational Psychology*, *96*(2), 195–203.