NORTHWESTERN UNIVERSITY

Middle School Students' Use of Epistemological Resources While Reasoning About Science Performance Tasks and Media Reports of Socioscientific Issues

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Learning Sciences

By

Brandy L. E. Buckingham

EVANSTON, ILLINOIS

December 2017

Abstract

Middle School Students' Use of Epistemological Resources in Reasoning During Science Performance Tasks and About Media Reports of Socioscientific Issues

By Brandy L. E. Buckingham

The goal of science education is to prepare students to make decisions about the complicated socioscientific issues that are an inescapable part of modern life, from personal medical decisions to evaluating a political candidate's environmental platform. We cannot expect adults to maintain a deep conceptual understanding of the current state of every branch of science that might prove relevant to their lives, so we must prepare them to rely on other knowledge to make these decisions. Epistemological beliefs about scientific knowledge - what it is, its purpose, how it is constructed - are one type of knowledge that could be brought to bear when evaluating scientific claims.

Complicating this situation is the fact that most adults will get most of their information about these socioscientific issues from the news media. Journalists do not have the same goals or norms as scientists, and this media lens can distort scientific issues.

This dissertation addresses the question of whether we can assess epistemological change in a way that gives us meaningful information about how people will apply their epistemological understanding of science when they make decisions in the real world. First, I designed a written assessment made up of performance tasks to assess middle school students' implicit epistemological beliefs, and looked at whether we can use such an assessment to see epistemological change over two

years. I then gave the same students news articles about whether there is a link between vaccines and autism and looked at their reasoning about this issue and how the journalistic features of two different articles impacted their reasoning. Finally, I examined the external validity of the epistemology assessment by looking at whether it predicted anything about students' responses to the news articles.

While I was able to find evidence of differences between eighth graders' and sixth graders' use of epistemological resources within the performance tasks, I found that their reasoning about the socioscientific issue was heavily dependent on the choices made by the journalists who wrote the news articles. The epistemological assessment gave us some information about their reasoning about the news articles, but this, too, was very highly dependent on the article itself. If our goal is to facilitate reasoning about socioscientific issues in the real world, we need to keep in mind the impact of the media on that process. These findings have implications both for how we teach science toward that goal and how we assess our progress.

Acknowledgments

This dissertation has been a very long process, much longer than average, and would not have been possible without the help and support of many people. First, the teachers and students of Maple and Park View, the two schools in which I did my research; obviously, none of this would have been possible without their cooperation.

I have said for a long time that there is no one I'd rather sit down and talk about these topics with than my advisor, Brian Reiser. He's excellent at asking the right questions at the right times, and has always encouraged me and insisted that my results are useful and interesting even when I was less than confident about them. I'm grateful that even as I've battled life and medical problems stalling my progress, he's never made me feel guilty for falling behind the typical schedule and never wavered in his confidence that I would finish.

My other committee members, Bruce Sherin and Reed Stevens, have provided feedback that will be invaluable as I turn this dissertation into publishable journal articles. They along with several other faculty members in this department have greatly influenced my theoretical and methodological perspectives.

As influential on my experience as a Learning Sciences graduate student and perspective on the field as any faculty member was the student-led Informalists research group, in particular founders Jessica Umphress, Danny Cohen, and Tiffany Simons-Chan. They were my main mentors in my early years in this program, and the Informalists, which I eventually took over running, became a central part of my LS experience.

My husband, Robbie Buckingham, has supported me in countless ways over the past decade (and before). As we've navigated the inherent complexity of being a dual-academic couple, including living in separate cities/states/countries for the first five years of this program, he's taken career risks (including the ultimate taboo, turning down a tenure-track job for a second postdoc) for the good of our marriage and family. Since our son was born, he's been not just an involved father who has taken on his share of parenting duties, but has been proactive in relieving me of those duties when I need to be, whether for work or mental health reasons. He's been a constant source of emotional support as I've dealt with my mental health problems, and doesn't hesitate to do whatever I'll find useful to minimize the impact of my executive dysfunction on my life and my work. I don't know how I would have made it through this without him.

While it's true that having a baby pushed my timeline back a bit, as the first year of a child's life is never an easy one to integrate with this kind of work, I'm so thankful that I had my son, Elias, when I did. He literally saved my life.

This dissertation would not have been possible without the help of a number of mental health professionals. I was diagnosed with bipolar disorder when I was nineteen and ADHD just a few years ago, and these neurological conditions have conspired for a long time to prevent my progress on this work. I don't remember the names of every single one who has contributed to my wellbeing in the past decade, but in particular Dr. Stephen Rush, Dr. Debjani Sinha, and Scott Ries have helped me turn things around over the past three years through a combination of medications, education, and therapy.

I've also enjoyed the support of countless friends I've met online, some of whose real names I know and some I only know by screen names. Some I've met in person, some I've only chatted with over IRC, Tumblr, or Twitter. They've been cheerleaders, helped me learn about myself, shared their own experiences, collaborated on projects, and given me a community.

Table of Contents

List of Tables and Figures

Chapter 1:

Introduction

> A man may have a good deal of cultivation, a good deal of information, correct information at that, about things, but if he has never made a first hand acquaintance at some point with scientific ways of dealing with a subject matter, he has no sure way of telling the difference between all wool knowledge and shoddy goods. He has no sure way of knowing when he is using his mental powers most capable and fruitfully. An ability to detect the genuine in our beliefs and ideas, the ability to control one's mind to its own best working is a very precious thing. Hence the rightful place of science in education is a fundamental one, and it is correspondingly important to see to it that methods of teaching are such as to fulfill its true purpose.
>
> - John Dewey, 1916, p. 4

In the same essay on Method in Science Teaching, Dewey notes, “We must remember that altho in school we are always treating pupils as embryonic scientists who somehow get interrupted and cut off before they get very far, the great mass of the pupils are never going to be scientific specialists. (p.8)” Over a century later, this is still true, and his observations on the goals and purpose of science education are possibly more relevant than ever. But while Dewey claimed that for the majority of students, an understanding of science is valuable for “the added meaning it gives to the usual occurences of their everyday surroundings and occupations (p. 8),” in modern times an understanding of science—or a lack thereof—has much more far-reaching implications for the average citizen.

Today, candidates’ environmental platforms are a key issue in every election; modern medicine offers increasingly complex treatments and sometimes conflicting results; politicians pass laws based on claims about women’s health that contradict what we know from research. Public understanding of science is a life-or-death issue for many people, and impacts the quality of life of many more.

We can't possibly expect everyone to graduate high school with a deep conceptual understanding of every field of science that might prove relevant to their lives. Even if we could, keeping up with the current state of every field would be an insurmountable task for the vast majority of adults. While a woman might have a basic understanding of how nutrients and other chemicals are passed from her body to that of her growing fetus, this won't be enough for her to judge exactly which medications are safe for her to take while pregnant and which are not. If we can't expect people to rely solely on content knowledge to make these decisions, what else can they use?

As an example, let us look at a recent article on CNN.com, "One-month sugar detox: A nutritionist explains how and why" (Drayer, 2017). This is not a topic that is likely to be life-or-death for many people, but it is certainly the kind of science-related claim that people might base daily decisions on.

Most of the article is made up of quotes from Brooke Alpert, a registered dietitian who wrote the book *The Sugar Detox: Lose the Sugar, Lose the Weight -- Look and Feel Great*, claiming that giving up all sugar for a month will boost your metabolism. This includes the statement, "We had over 80 testers from all over the country, and they lost anywhere between 5 to 20 pounds during the 31 days, depending on their weight or sugar addiction," though no further information is given about this "test" or its methodology.

The article also contains some quotes from NAME, a professor of pediatrics at PLACE who has conducted peer-reviewed research related to this topic. The article claims, "Lustig's research revealed that when obese children eliminated added sugars from their diets for just nine days, every aspect of their metabolic health improved -- despite no changes in body weight or total calories consumed." It does not indicate that Lustig is in any way connectd to Alpert, her book, or her research.

Most adults who read this article will not have a detailed understanding of the biochemistry of metabolism and how we metabolize sugars, so they will need to rely on other knowledge and beliefs to decide whether the article's claims are valid and relevant to their lives. These might include things like whether the fact that Alpert wrote a book increases her credibility, or if it decreases it either because she's trying to sell a book or because the research wasn't peer-reviewed; whether they believe that Lustig's research on obese children is likely to be generalizable to themselves; whether the summary of Alpert's test of 80 subjects makes them believe that it was a credible study or if they need more information on methodology to make that judgment; or even what they believe were the goals and scientific expertise of the journalist who wrote the article.

These can be collectively referred to as epistemological beliefs: beliefs about what knowledge is, where it comes from, and how to evaluate it, and as Dewey and many others since his time have noted, an epistemological understanding of scientific knowledge can be valuable in this situation. This might include ideas about what scientific knowledge is, what processes are used to create it, what its purpose is, and what justifications should be used to support it. If the goal of science education is to prepare people to use their scientific knowledge in their everyday decision-making, an understanding of these issues needs to be a part of that goal.

The central question of this dissertation addresses this goal; specifically, how we can measure our progress toward this goal in a meaningful way:

How can we assess epistemological change in a way that gives us meaningful information about how people will reason about science-related situations in the real world?

This question has three parts, and the three papers that compose this dissertation each address one of these.

The first paper addresses the question of how we can assess epistemological change at all. In order to do so, we need to know what it is we're measuring, so I first look at theoretical approaches to epistemological beliefs and explain the one that I will be using in the dissertation, the idea of *epistemological resources*. I discuss the methodological consequences of this framework, leading to the main research question of this paper: What differences in epistemological beliefs over time can we capture with a written assessment? I then explain the assessment design features my theoretical framework led me to and how I translated them into written performance tasks which engaged students in scientific practices. I gave these assessments to middle schoolers in the fall of sixth grade and spring of eighth grade. I show how their answers to these free-response questions show evidence that the eighth graders are using the epistemological resources of *evidence* and *cause and effect* more than the sixth graders, who are using *correctness* and *amount of information* more than the eighth graders. This gives us evidence that we can measure epistemological change as defined by this perspective using this type of written assessment.

The second paper addresses the end of the central question: How do people reason about science-related situations in the real world? I show that, overwhelmingly, when people need to make decisions about these issues they tend to get their information from the news media; furthermore, these media accounts of scientific topics differ greatly from academic science texts. These differences can have a measurable impact on people's reasoning. If we want to know how people will reason about these situations in the real world, we need to look at how they reason when the information is filtered through this media lens. The specific part of this problem that is addressed by this paper is then: How do journalists' choices affect students' reasoning about a socioscientific issue from a news article? I selected two news articles about the purported link between vaccines and autism that varied in systematic ways that, theoretically, could influence their readers. Using the same population as in the first study, I gave each student one of the two articles to read, followed by open-ended questions and tasks. I found that the journalistic choices in these articles had a powerful effect on the reasoning of the students who read them; one of the articles set up a false equivalence between scientific research and other types of justification, which led to those students equating the two. Students who read the other article, which did not do this, did not show a relationship between how they justified their reasoning and what information they categorized as scientific evidence.

The third paper then brings these two studies together, addressing the heart of the question: now that we have measured epistemological beliefs and change, and looked at how those same people reason about socioscientific issues in the news, does the first tell us anything about the second? To answer this, I used scores from the epistemology assessment as predictors in binary regressions on behaviors in the news assessment. I found that, as I saw in the second study, the strongest predictor overall was which article the students read. But there was also some information from the epistemology assessment that predicted behaviors on the news assessment. For example, students who valued research data or who talked about cause and effect more were more likely to focus on the research in the articles and less likely to focus on the anecdotal data in the first article.

Taken together, these results imply some hopeful conclusions, but also highlight some additional issues for the field to investigate. This is a promising start toward the goal of assessing epistemological change in a way that gives us meaningful information about how people will behave outside of school. It shows that we can measure this on a large scale, and that if we find the right indicators to look at, we can use it to predict how people will reason in other situations. However, the large disparity in reasoning between the students who read the two articles indicates that this is only one part of the larger problem of helping students use science to make decisions in their daily lives. We need to give more attention to issues of media literacy and how they interact with science literacy when people encounter these situations.

Chapter 2:

Assessing Epistemological Change Via Written Performance Tasks

The Framework for K-12 Science Education and the Next Generation Science Standards (NGSS) make it clear that scientific literacy is not only about getting the correct answers on a test. They explicitly state that the goal of scientific literacy is to help learners navigate situations in the real world: "[S]ome knowledge of science and engineering is required to engage with the major public policy issues of today as well as to make informed everyday decisions, such as selecting among alternative medical treatments or determining how to invest public funds for water supply options (NRC, 2012, 3)." "When comprehending current events, choosing and using technology, or making informed decisions about one's health care, understanding science is key. … All students no matter what their future education and career path must have a solid K–12 science education in order to be prepared for college, careers, and citizenship (NGSS Lead States, 2013b, 3)."

The idea that average people, not only scientists, should be scientifically literate is not a new one, nor is the idea that this literacy goes further than simple knowledge of facts. In his 1916 speech to the National Education Association, Dewey stated that the goal of science education should be scientific thinking, and that "science is not only knowledge, but it is knowledge at its best, knowledge in its tested and surest form… if [a man] has never made a first hand acquaintance at some point with scientific ways of dealing with a subject matter, he has no sure way of telling the difference between all wool knowledge and shoddy goods. He has no sure way of knowing when he is using his mental powers most capable and fruitfully. (pp 3-4)"

Over time, a variety of models of scientific literacy have emerged (Shen, 1975; AAAS, 1985; Miller, 1998; NRC, 2006; 2009; OECD, 2006). All of these models agree with Dewey that in order to be scientifically literate, one must not only know and understand the current products of the scientific enterprise, but also have some level of understanding about how that information came to be. Though the phrasing varies, they all claim that a scientifically literate individual must have some grasp of the *nature of scientific knowledge* and of *how science is practiced in the real world in order to construct this knowledge*.

These ideas are reflected in the NGSS in the form of eight science and engineering practices in which students should be able to engage, which are "not only intended to strengthen students' skills in these practices but also to develop students' understanding of the nature of science and engineering (NGSS Lead States, 2013, xx)." The listed practices are:

1. Asking questions and defining problems
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations and designing solutions
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information (NGSS Lead States, 2013, xx)

These are engaged in through an ongoing cycle of creation, evaluation, and revision of scientific knowledge products such as explanations, arguments, and models. As Dewey alluded to, not every student will grow up to be a scientist and need to engage in all eight of these practices through all three parts of the cycle regularly. But, at a bare minimum, all adults in our society need to be able to evaluate different forms of scientific claims in order to make decisions in everyday life. The NGSS suggests that reflective engagement in these practices, supported by a conceptual understanding of current scientific knowledge, will allow students to develop the epistemological understanding of science that they need in order to carry out this day-to-day decision-making

Does the model of science education outlined in the NGSS actually change students' epistemic knowledge in a way that furthers the goal of preparing them to apply this information in the real world? Currently, there is little data with which to answer this question. Research on how to best encourage epistemological change in science is limited, with sometimes conflicting results. Many of these conflicts stem from disagreements on how best to measure this change, which in turn stems from deeper theoretical differences.

Theories of epistemological beliefs

On a general level, most theorists agree that people have some kind of knowledge or beliefs about what scientific knowledge is, how it is created, and how it is used, and that this knowledge guides them when they are engaging in a variety of everyday activities. For example, a woman who reads a newspaper article about a scientific discovery has to have some way of deciding whether she believes the scientists' claims. She may see scientists as arbiters of truth and believe whatever they say, or she may have a more nuanced understanding of the process of scientific knowledge creation and see possible flaws that make her question the article's claims. The second option requires not just an understanding of scientific methodology, but also the belief that scientific knowledge is tentative and can be questioned, and that there are criteria for evaluating a scientific claim and the evidence for that claim.

How one assesses epistemological beliefs and change depends on how one conceptualizes these beliefs in the first place. There have been several schools of thought on epistemological change in the past two decades, and each would have different implications for measuring the impact of engagement in scientific practices. Theoretical perspectives have largely varied along two axes: whether they characterize this knowledge as static across contexts and knowledge domains or context-dependent, and whether they view this knowledge as explicit or implicit. Both of these factors impact the way that researchers have chosen to assess this knowledge. Table 1 shows four theoretical perspectives on epistemological beliefs and their placement along these axes.

Table 1: Four theoretical perspectives on epistemological beliefs

Developmental	Dimensional	Nature of Science	Epistemological Resources
Domain-general, mostly static across contexts	Domain-general but varying along multiple dimensions, static across contexts	Domain-specific but static across contexts	Entirely dependent on context
Explicit	Explicit	Mostly explicit	Implicit

Developmental

King & Kitchener's (2004) Reflective Judgment Model (RJM) is one that looks at epistemological change developmentally. These researchers often focus on very generally-worded beliefs that cut across many domains, such as "Knowledge is assumed to exist absolutely and concretely" or "Beliefs are justified within a particular context by means of the rules of inquiry for that context and by context-specific interpretations of evidence (p. 7)" and group responses into distinct stages, reflecting their belief that epistemological beliefs are stable entities that can be stated explicitly. Although they mention the possibility that the amount of support in the assessment context may affect a person's performance, reflecting different functional and optimal stages of development, they do not explore the effects of context. They view epistemological development as something that can only be measured on a large scale, over the course of years. In the RJM, students do not even reach Stage 3 out of 7, on average, before graduating high school. Such a model is not useful when looking at changes that might take place over the course of one year of elementary school, assuming you believe that such changes are possible.

Multi-Dimensional

Others have broken down personal epistemology into multiple dimensions to be assessed separately, such as the structure of knowledge versus the stability of knowledge (e.g. Schommer-Aikins, 2002; Jehng, Johnson, & Anderson, 1993). Most of these researchers have relied on Likert scales or similar to assess how strongly people believe statements about the nature of knowledge, or sometimes assessments that ask them to state their own beliefs (Duell & Schommer-Aikins, 2001). Both of these types of assessment reflect a belief that people are consciously aware of their own beliefs and can verbalize them accurately. It also assumes that the beliefs that come to mind when a person is taking such a written test are the same ones that the person would call on in other contexts. Despite the use of multiple dimensions, these researchers often see these as general beliefs that cut across many or all domains of knowledge.

Nature of Science

Many researchers now rely on theories that assume that beliefs about knowledge may vary depending on the domain of the knowledge. According these theories, to determine the impact of engagement in scientific practices we should look specifically at changes in beliefs about scientific knowledge and its construction. Even here, though, there are differences in the conceptualization of these beliefs that have significant repercussions for their assessment.

As in more domain-general multidimensional epistemology research, some researchers rely on written questions that assume that epistemological beliefs are explicit, stable, and can be verbalized, such as the Views of the Nature of Science (VNOS) questionnaire (e.g. Schommer-Aikins & Duell, 2013, Khishfe, 2012). These researchers have found that simply engaging in the practice of science is not enough to change these explicitly-stated beliefs, and argue instead that instead explicit instruction is required (Khishfe & Abd-El-Khalick, 2002; Khishfe, 2008).

Epistemological Resources

There is evidence that epistemological beliefs can be much more context-sensitive than simply varying by domain, and that learners can make use of epistemological ideas that do not align with those they can explicitly verbalize. Salter & Atkins (2013), for example, found that while students' VNOS answers were consistent over the course of a semester, their performance during classroom activities showed increased sophistication at the end of the semester.

To account for such inconsistencies, Hammer and Elby (2002) have proposed a theory of personal epistemology that conceptualizes it as manifold, context-dependent, and often implicit rather than unitary, stable, and explicit. What we might call an epistemological belief is constructed as needed within a context from simpler epistemological resources (ERs). The same person may make use of different ERs in different situations, and may not be able to state any explicit "belief" that would encompass these uses. This would explain why, for example, some people perform tasks as though they hold more sophisticated epistemological beliefs than they are able to state explicitly on assessments, as though they hold more sophisticated beliefs in one situation than another.

Sometimes, people do behave as though they hold consistent beliefs across similar contexts. ER theorists have incorporated the idea of "framing" to account for this (Hammer, et al., 2004;

Scherr & Hammer, 2009). When someone activates the same locally coherent set of ERs repeatedly in similar contexts, it becomes increasingly likely that they will activate those ERs as a set in other contexts that they judge to be similar. This set of ERs becomes known as a "frame," and can behave as a larger cognitive unit. The process of judging what a context is similar to in order to activate an appropriate frame is "framing." Hammer, et al. demonstrated how this can account for shifts in students' thinking: when they first see a situation as similar to one context but then decide (consciously or subconsciously) that it is similar to another, this change causes a shift in which frame they are activating. Scherr and Hammer observed students working in groups on a physics worksheet, and found that the group behavior indicated sharp shifts as students switched from framing the activity as *filling in the worksheet* (activating ERs for concepts like *authority as a source of knowledge*) to *discussing ideas* (activating ERs like *fellow students as sources of knowledge*). These frames appeared to be fairly consistent between groups of students.

Like the other theories of personal epistemology, this theory has methodological consequences. Asking learners to explicitly state their epistemological beliefs is unlikely to provide a complete picture if they are not consciously aware of which ERs they are using at a given time. In addition, any assessment can only be a snapshot of what ERs the learner is making use of at that time and in that context. Because of these considerations, the most common method of assessment in this research is to observe learners completing actual tasks and to analyze how they frame those tasks epistemologically while they engage in them (Louca, Elby, Hammer & Kagey, 2004; Rosenberg, Hammer, & Phelan, 2006; Berland & Crucet, 2016). For example, Louca, et. al. (2004) show examples of a third grade class discussion of leaves changing color. At first, the students only give teleological explanations for the phenomenon, but the teacher provides examples of teleological vs.

mechanistic explanations that they are likely to have encountered in their everyday lives. Luca et. al. demonstrate that a resources perspective, which would say that the teacher helped them reframe the activity and activate their *mechanism* ER by showing them other activities they already frame that way, accounts for this episode better than developmental or explicit-beliefs framework.

The goal in these situations is not to put a label on the learner specifying exactly what they believe, or what level or stage they are in, but to see whether they are making use of productive epistemological frames in appropriate contexts and which contextual cues lead to their shifting their epistemological frame.

Approaching the idea of epistemological change from an ER perspective, the original question can be restated as: Does the approach to science education outlined in the NGSS help students develop productive epistemological frames and apply them appropriately?

Epistemological Resources in the Literature

When initially proposing their theory of epistemological resources, Hammer and Elby (2002) suggested a wide variety of possible ERs, focusing on domain-general resources that could be brought to bear on understanding the nature of any type of knowledge. These included ERs concerning where knowledge comes from such as *knowledge as propagated stuff, knowledge as free creation,* and *knowledge as fabricated stuff;* ERs concerning epistemological activities such as *accumulation*, *formation,* and *checking*; and ERs for forms of knowledge such as *stories, rules,* and *facts.*

While these are applicable to all knowledge, including scientific knowledge, subsequent researchers have also looked at ERs that are more specifically relevant to understanding what scientific knowledge is and how it is created. Jiménez-Aleixandre (2014) suggests three ERs relevant to genetics learning: *probabilistic thinking, acknowledgement of uncertainty,* and *commitment to consistency of explanations*, all of which are also applicable to other branches of science. van Aalst's (2012) analysis of learners' notes while learning about states of matter in a computer-supported environment suggested that they activated ERs such as *questions, prior knowledge, theories, new information,* and *puzzles.*

There are some ERs that have been studied repeatedly with respect to science learning.

Mechanism

The first of these is perhaps the most-studied ER, though the wording varies: *cause and effect*. Louca et al. (2004) look at third graders' use of *mechanistic* versus *teleological* explanations, while Rosenberg, Hammer, & Phelan (2006) use epistemological framing to account for students' perceptions of science as made up of *isolated facts* versus *causal chains*. Russ et al. (2009) argue for greater focus on *mechanism* in classroom assessment over *textbook correctness*. In defending their choice to specify *mechanism*, they note that "Causal reasoning serves as a starting point for the pursuit of underlying mechanistic explanations, but causal reasoning alone does not define it (p. 7)." In other words, stating that one event causes another event is not the same as giving a full account of *how* this causation happens.

Evidence

The use of *evidence*, of course, is a much-covered topic in the science education literature, and has been studied from an epistemological resources perspective. Berland and Hammer (2012) show how students are more likely to call on evidence when their discussion is framed as an argument as opposed to an idea-sharing session. In addition, when the teacher wants to reframe the discussion away from argumentation and more toward instruction, he explicitly removes the graph students were using as their evidence, signaling that they were no longer supposed to refer to evidence because they were no longer constructing scientific arguments. Sandoval and Cam (2011) found that, when evaluating scientific arguments, children valued evidence over causal mechanism or calls to authority.

Correctness

Hutchison and Hammer (2009) and Russ et al. (2009) both show how a teacher's focus on *correctness*, in the form of an answer matching that given in a textbook, can get in the way of learners' use of *mechanism*. Berland and Crucet (2016) make use of a slightly different type of *correctness*—literal resemblance to objects. Elby (2000) referred to this ER (or knowledge element, as this paper was a direct precursor to the epistemological resources framework) as *what you see is what you get*. Delgado and Lucero (2015) also identified *what you see is what you get* when studying the ERs that students activated when studying scales and graphing.

Impact of Epistemological Resources Theory on Assessment

If our true concern is how students will behave "When comprehending current events, choosing and using technology, or making informed decisions about one’s health care (NGSS Lead States, 2013b, 3)," and we accept a view of epistemological beliefs as context-dependent, context is an essential part of an assessment. If students are likely to apply different frames when participating in scientific practices than they do when answering questions asking them to state their beliefs explicitly, then the latter is not likely to give us useful information about how they will behave in the real world.

Analyses of video data and classroom discourse, though they are likely the most valid way to assess students' framing, are time- and labor-intensive and require much training to execute. These methods would rarely be feasible to conduct research on large numbers of students, nor would they be convenient if we want teachers, school districts, or even states to be able to assess epistemological change across grade levels for dozens to thousands of students. For these purposes, written assessments would be ideal. But, as we've already seen, simply asking students direct questions about their beliefs is problematic if you view those beliefs as implicit and context-dependent. Instead, any written assessment would need to give a student the opportunity to make use of their epistemological beliefs within the context of an appropriate task.

Written performance-based assessments have been used to assess conceptual knowledge in science for a long time (e.g. Adams & Callhan, 1995; Schneider, Krajcik, Marx, & Soloway, 2002). Their use in assessing epistemological beliefs, however, has been limited. The VNOS-C contextualizes some of its questions in terms of having learners analyze scientific thinking or arguments, but the actual questions it asks still focus on the explicit statement of beliefs (Abd-El-Khalick, Lederman, Bell, & Schwartz, 2001).

Written assessments have only occasionally been used for epistemological assessment from an ER perspective. Hammer & Elby (2003) used written essays on physics topics to assess students' developing epistemological framing. These presented students with a physics problem, such as where a dropped quarter will land, and asked them to not only defend their own answer, but to think of another answer someone might give, why they might give it, and refute the other answer. They were able to infer that students activated ERs related to concepts such as prior knowledge and reconciling inconsistencies, without the students explicitly discussing these concepts in their essays. Sandoval and Cam (2011) gave students four different stories about people debating scientific explanations, asking them to evaluate the reasoning used in these stories. Based on students' decisions and reasoning, they theorized that the students were drawing on epistemological resources such as *claims need evidence* and *causal mechanisms must be plausible*. While in this study, they interviewed the students, a similar task involving evaluating a story about others' reasoning could easily be adapted to a written assessment. And van Aalst (2012) looked at students' notes and answers during a computer-based unit to infer the ERs that students activated while learning about states of matter.

Performance-based assessments could engage students in the creation and evaluation of scientific knowledge products, in the context of the same practices in which they have engaged in the classroom. However, there are several possible limitations that must be taken into account and studied. Will reading about a situation activate the same ERs that participating in a similar situation would, or a close enough approximation? Can such a written assessment capture changes in epistemological framing over time?

The goal of this research is discover what epistemological changes are possible when students are taught using reform-based science curricula, using a performance-based written assessment.

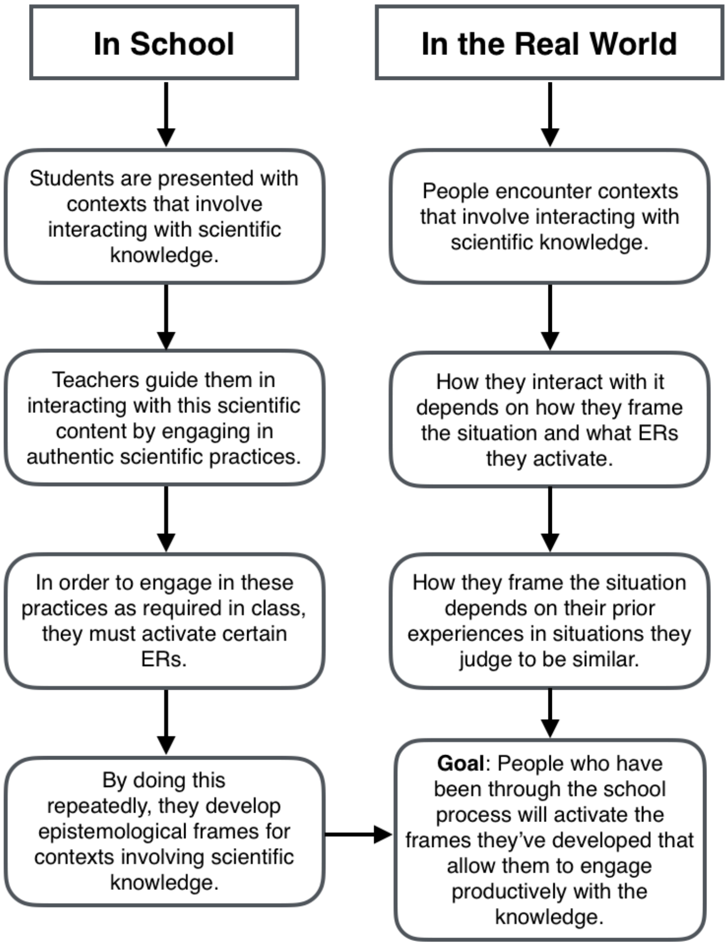
Theoretical Framework

The design of this study is based on a theoretical framework that assumes that students call on ERs to engage in all types of classroom activities, and that they take into account context cues, from the type of task to instructions from their teacher, when unconsciously deciding how to frame an activity. Epistemological learning, then, is a change in the way the students frame activities epistemologically—the development of new frames with more productive ERs and the ability to call on them when faced with activities for which they would be appropriate in the future.

The Frameworks and NGSS assume that engaging with scientific content knowledge via reflective and authentic scientific practice leads to more productive epistemological beliefs. Our theoretical framework suggests that the process of engaging in authentic scientific practice in the classroom, under the guidance of teachers who will help students develop productive epistemological frames to carry out this engagement, will result in the students' developing epistemological frames that they will then call on when faced with appropriate situations outside of the classroom.

This theoretical framework, a summary of which can be seen in Figure 1, means that if we are interested in epistemological change, we must look at shifts in the way students frame contexts, and thus which ERs they activate in order to engage with those contexts. Therefore, to measure change we need to assess what ERs they are bringing to bear on relevant contexts.

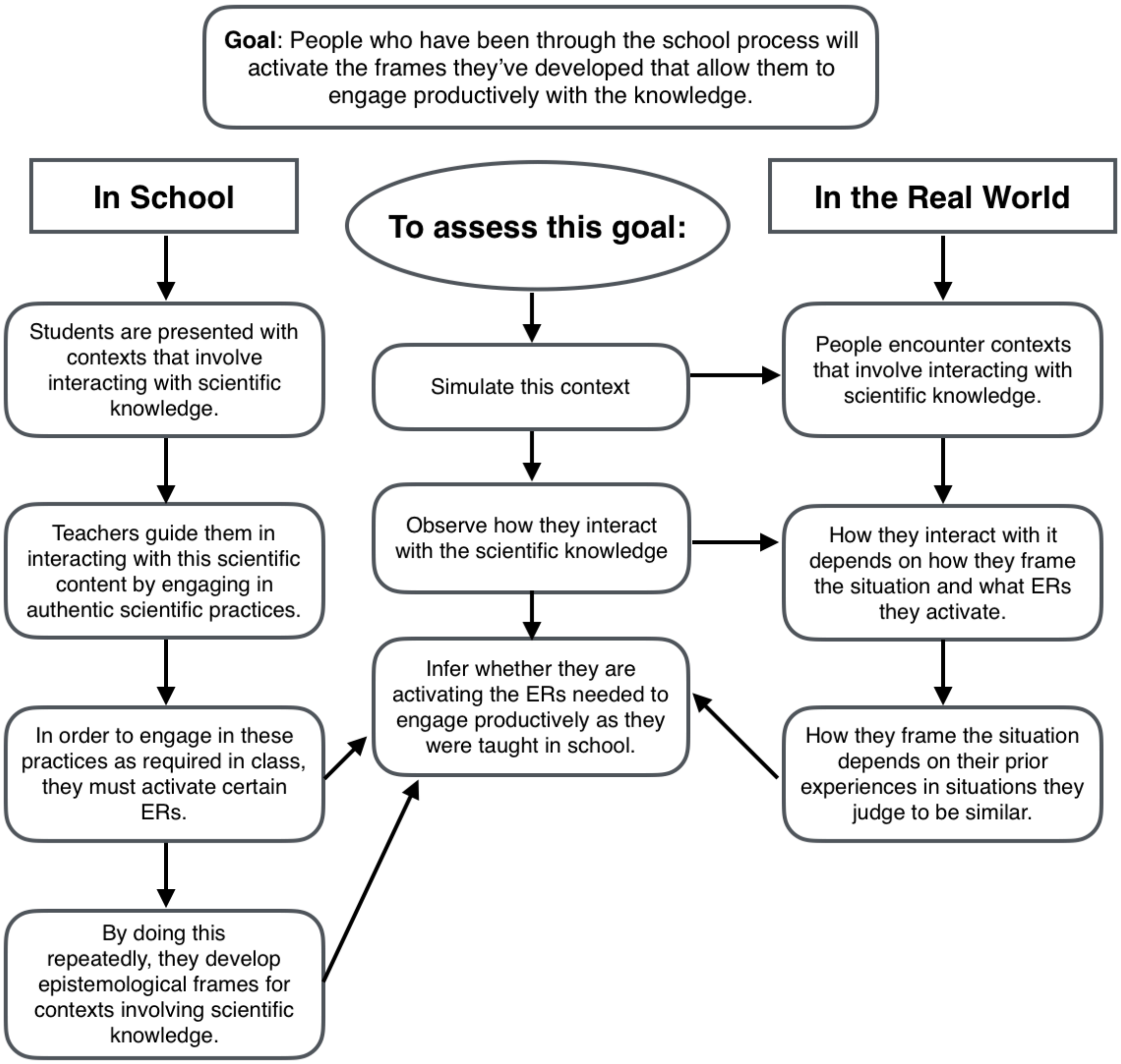
Figure 1: Model of how epistemological change in school can impact behavior outside of school



However, according to this theoretical perspective, an assessment can only reveal how a student is framing the assessment itself. Therefore, assessments must mirror the real world situations in which we hope students will activate productive epistemological frames as closely as possible.

In this case, we hope that students will apply the frames they developed while engaging in scientific practices in the classroom in order to engage productively with situations outside of the classroom. Therefore, we need the assessment to provide a context in which it would be appropriate to make use of these practices, to see how students frame those situations epistemologically. The implications of our theoretical framework on assessment can be seen in Figure 2.

Figure 2: Impact of this model on assessment.



This theoretical framework led us to create written performance tasks focused primarily on the evaluation of scientific knowledge products. What aspects of an explanation, argument, or model a student looks for in an evaluation can tell us what ERs they believe are important in that situation, much like Sandoval and Cam (2011) were able to infer whether students valued a mechanistic explanation over evidence to support a claim.

So, for example, if we present students with an argument and ask them what the argument's strengths and weaknesses are, they may call on ideas like evidence, mechanism, accuracy, prior knowledge, or the word of authority figures. If a person in the real world is reading a scientific argument in the news, we would hope that they would look for evidence and possibly a plausible mechanism. We would hope that they would not simply accept an argument because it came from an authority figure, or because it "sounds right" to them, especially if the competing argument could offer better evidence and a more plausible mechanism. Different people will activate different ERs to evaluate this argument, some more productive than others, and those will be evident in their evaluations.

I will go into greater detail about how I have applied this theoretical framework to the design of our assessment in the Methods section below.

Methods

Research Questions

The overarching research question for this study is *How do students' epistemological beliefs change over time?* As outlined in the theoretical framework, measuring epistemological change involves looking for evidence of the use of specific epistemological resources when doing work in a relevant context. I believe that two to three years' experience engaging in authentic, reflective scientific practices in the IQWST curriculum will help the eighth graders to call on more appropriate epistemological resources than the sixth graders. Specifically, I hypothesize that the eighth graders will call on Evidence and Cause and Effect more often than sixth graders, while calling on Correctness and Amount of Information less often than sixth graders.

Participants

Participants in this study were 180 students from ten sixth grade classes and 98 eighth grade students from eleven eighth grade classes at two midwestern middle schools (one sixth grade and one eighth grade teacher in each school). 138 students were males and 138 were females, with two students not indicating their gender. This is a cross-sectional study, not longitudinal; all data was collected during the 2010-2011 school year.

Both middle schools use the IQWST curriculum (Krajcik, et al., 2004) for their sixth through eighth grade science classes. This curriculum focuses on teaching science content through engagement with authentic scientific practices, with particular emphasis on explanation, argumentation, and modeling, including reflection on the connections between practice and content. At one school, the teachers have several years' experience both using IQWST and helping to train other teachers to use it, while 2010-2011 was the first year the other school was implementing the curriculum; therefore, eighth graders at one school have been through two previous years of IQWST-based science classes while eighth graders at the second school have not.

School using the IQWST curriculum were chosen because I want to see not only that epistemological change can be measured over this two-year time frame, but that it can be seen specifically when students have been engaging in authentic scientific practices and reflection as suggested by NGSS.

Design of the Assessment

Every scientific practice is made up of a cycle: creating, evaluating, and revising. Scientists need to be able to do all of these, whether it is when they are asking questions, carrying out investigations, or communicating their results. However, most students will not grow up to become scientists. The goal of science education is not to prepare every student to work as a scientist; it is to prepare them to deal with science that they are likely to encounter in their daily lives. They are unlikely to need to actually create scientific investigations or models; however, they are very likely to need to evaluate scientific claims, methods, evidence, and reasoning. Therefore, similar to Sandoval and Cam (2011), our assessment focuses on this aspect of the practices more than creation or revision, although it does touch on all of them. But I am most concerned about how students will evaluate scientific knowledge claims they come across, in order to use scientific knowledge in their future decision-making.

Unfortunately, a written assessment given in school will always carry with it connotations that real-world scientific situations do not carry. Students know that they are being tested; even when told there is no right answer, they may apply frames that center on providing the correct answer or looking good under evaluation. This is one drawback of this type of assessment as compared to observation of real-world behavior, but it is a necessary tradeoff when hoping to evaluate a large number of learners.

In designing a set of performance tasks that will give us the most complete picture of the ERs learners call on when engaging in scientific practices, I had several priorities to keep in mind. Our assessment was based on six design principles:

1. The situations provided need to resemble situations the learner may encounter in the real world and the tasks must ask learners to engage in practices that would be relevant to that situation.

2. The questions asked about the provided situations must not ask the learner directly or explicitly about their ERs. Rather, they should prompt learners to reveal their thought processes so that we can see which ERs they are actually making use of.

3. Learners need to be able to express their thinking freely. If the questions are too tightly constrained, we may not get a complete enough picture of their thinking.

4. When possible, learners should be given multiple opportunities to express their thinking about each situation. They may not talk about every aspect of the situation that they are considering in a single answer.

5. When possible, any scientific content should use concepts the learner is familiar with. Research has shown (Schwarz et al., 2009) that learners are better able to engage productively in scientific practices when they are familiar with the content.

6. The situations provided should give learners the opportunity to make use of the ERs that we believe would be productive when faced with a similar situation in real life. For example, not every instance of engaging productively in a scientific practice requires the learner to think about evidence, so we need to be sure we include at least one opportunity for them to do so if that is something we are interested in.

Our focus on evaluation of scientific knowledge products helps us to design an assessment that implements all six of these principles. First, as we demonstrated above, it allows us to implement principle #1, resembling situations in the real world, more closely than focusing on other activities such as designing an experiment would. Second, it allows the questions to be open-ended (principle #3) and not have to ask about the ERs directly (principle #2) while still being able to vary the situations they are evaluating in order to address particular ERs (principle #6). We can give them situations that vary in terms of which ERs an argument/explanation/model was based on to find out which ones they find most valuable; because our current goal is to find out what is *possible*, rather than what is most common, this scaffolding gives us a better chance of seeing students at their highest potential. Given all of this, principles #4 and #5 are then easy to craft such an assessment around.

With these priorities in mind, we created four performance tasks (the full assessment can be found in Appendix A):

Question 1 presents the learner with data, in both table and graph form to minimize the chances that misunderstanding the data would affect the learner's answers. This data shows how many pills each subject took and how long the subject's headache lasted. On average, the people who took more pills had shorter headaches. It then relates an argument between two scientists, Peter and Anne. Peter claims that the pills do not work, and cherry picks a few data points to support this argument. Anne claims that the pill does work, but she does not refer to the data provided; instead, she says she talked to one man who took the pill and he said that it helped his headache.

Learners were first asked whether they believed the pill worked, and why. We were curious as to whether they would use the data themselves to form an opinion, or the arguments from the scientists. We then asked them to list the strengths and weaknesses of each person's argument, and then to say, regardless of whether the arguments were correct or not, who did a better job of arguing for their answer. We wanted to see if learners would mention the scientists' use (or non-use, or misuse) of the data or the fact that neither argument mentions a causal mechanism to explain why the pills would or would not work.

Question 2 presents the learner with a scenario in which a classroom janitor unplugged a fish tank, causing some of the fish to die over the weekend, as well as five possible explanations for why the fish died. We then asked them to circle the best explanation, and next to each one write why it is or is not the best explanation. We attempted to discourage learners from relying on correctness by ensuring that none of the five explanations contradict each other. However, as some explanations give more information than others, they could believe this extra information to be incorrect even if they believe that the information common to all explanations is correct.

Statement A is simply a claim: 90 degrees is too hot for fish to live. Statement B gives a full causal chain to account for why fish cannot survive in warmer water. Statement C provides one link in this causal chain, but not the full explanation. Statement D includes all of the information in B, but also the extraneous statement that the surviving fish were probably hungry. Statement E is approximately the same length as B and D, but is made up of the claim in A repeated with different wordings. and no causal statements explaining it at all.

These were designed to gauge how much a learner is relying on Cause and Effect in evaluating an explanation. If a learner is evaluating these statements primarily based on their use of Cause and Effect, they should find B the most appealing, as it contains a full causal chain and nothing else. If they are calling on the idea of Cause and Effect, but also evaluating them based on the idea that more information is always better, they may choose D. If they are primarily looking at length, it will be a tossup between B, D, and E.

Question 3 is the only one that asks the learner to construct a scientific explanation rather than evaluate a scientific knowledge product. It presents them with three scientific phenomena to choose from: water disappearing from a dish, carnivores moving into an area resulting in the plant population going up, and smelling a perfume from across the room. We chose these because, given the curriculum used in this school district, the sixth graders should be familiar with at least two of the three and the eighth graders with all three, reducing the chance that lack of content knowledge would interfere with their answers. The learners are to choose one of these three phenomena and write an scientific explanation for it. This item was included to see to what extent the learners value Cause and Effect when constructing a scientific explanation.

Question 4 asks the learner to evaluate two scientific models of condensation. Both show a container of water with plastic wrap over the top. One model is focused on showing the water droplets in the air, and also shows the water at the bottom of the container and on the wrap as particles. The other is focused on showing that the droplets are moving in the air and making their way to the plastic wrap, but shows the water at the bottom and top as solid blue. They are asked to state each model's strengths and weaknesses, tell which model is better and why, and explain what they would change about each model and why. Another subquestion asks whether the models are generalizable to another phenomenon, but this is not included in the present analysis. This item was designed to determine to what extent learners rely on the models' literal resemblance to reality versus their portrayal of a causal mechanism explaining the phenomenon. For example, a learner who prefers the second model because it shows the water as a blue solid is paying more attention to its resemblance to what she would see in reality, whereas a learner who prefers it because it shows that the water particles move up to the plastic is more concerned with explaining the phenomenon.

Coding

We are interested in finding out *what epistemological changes are possible* over the course of two and a half years of school, meaning that in an initial analysis we want to cast as wide a net as possible. In addition, we want to minimize the chances that sixth graders' thinking simply seems less sophisticated because they are not as skilled at explaining their thinking as eighth graders. Therefore, rather than taking a more detailed look at *how* students make use of particular ERs, we are simply going to look at *what* ERs they use.

Our first three codes are based on ERs we previously discussed that have been identified multiple times in the science education literature: *cause and effect, evidence,* and *correctness*.

Cause and Effect. While most studies in the literature specifically look at *mechanism*, we wanted to cast a wider net. We know that sixth graders may not have the vocabulary needed to articulate their thoughts well enough for us to definitely say whether they're talking about *mechanism* specifically, which might artificially inflate the difference between the sixth and eighth graders. Because of this, we are looking at a more basic concept, *cause and effect*—any indication that the student was taking into account the way one event leads to another or why/how things happen. Certainly, future research will want to tease apart how to capture learners' development toward true mechanistic thinking, but right now we are going to look at learners' reliance on the idea of *cause and effect* in a more general sense.

Evidence. Again, we want to cast a wide net as far as what to include as *evidence*. Rather than a more detailed look at what the students value about the evidence, we apply this code anytime they take into account anything about the type, content, quantity, or quality of scientific evidence in their answers.

Correctness. It may seem counterintuitive at first to categorize *correctness* as a "less productive" epistemological resource, as we do for this analysis. Correctness is, of course, valued in science; we want data that accurately represents reality, and we want scientists to draw the correct conclusions from this data in order to form theories that are also correct. In science, though, we assess correctness in terms of concepts such as those above—a theory is "correct" if it explains available evidence and provides sufficient causal mechanism, and this "correctness" is a tentative state that may change if the evidence changes or another theory provides a more plausible mechanism or accounts for more evidence. We are using *correctness* in both of the ways it has been used in the literature: when a student relies either on factual correctness or literal similarity, without taking into account mechanism or scientific evidence.

Amount of Information. In developmental theories of epistemology, whether domain-general or specific to science, the lowest levels often involve students viewing knowledge as being made up of a list of discrete facts rather than as interconnected theories (Carey and Smith, 1993). In Schommer's (1993) multidimensional epistemological questionnaire, one of the dimensions is *Simple-Unambiguous Knowledge* and contains ideas such as Seek Single Answers and Avoid Integration (of facts). Schommer found that junior college and university students whose parents were more educated and/or encouraged independence when they were growing up were less likely to see knowledge as simple and unambiguous. De Vow, Bulte, and Pilot (2002) noted this as a major flaw in chemistry education, historically—that textbooks tend to present chemistry as a long list of context-free facts instead of interconnected theories grounded in real world experiences.

In one study looking at epistemological resources and framing, Rosenberg, Hammer, and Phelan (2006) describe instances of learners treating scientific knowledge as something "comprised of isolated, simple pieces of information (p. 270)." These learners focus on vocabulary terms over conceptual understanding, and view getting the right number of facts written down, in roughly the right order, as more important than thinking about the causal connections between the facts. Their framing of the assignment shifts, though, and they begin to focus on their conceptual understanding of the rock cycle and on telling a cause-and-effect story.

This idea of knowledge as a list of uncomplicated, isolated facts came up repeatedly in students' responses to our written performance tasks. We chose to call the ER *amount of information* because, most often, it involves students valuing the number of details or length of writing over the content of the writing, such as mechanism or evidence.

Along with *correctness*, we view this as an ER that may be applicable in many scientific situations, but if it is not paired with some concern for the content of the scientific knowledge—in this case, specifically, whether it contains information about *cause and effect* or *evidence*—it cannot be used as productively as when the content is taken into account.

Questions 1, 2, and 4 were coded for four different ERs, although not all of the codes were applicable to all questions: Correctness, Amount of Information, Cause and Effect, and Evidence. Table 2 gives a summary of the meanings of these codes.

Table 2: Codes

Code	Indicates that answer mentions or relies on…
Correctness	Factual correctness of information and/or literal similarity to reality, without considering evidence or mechanism.
Amount of Information	Length, number of details or otherwise the amount of information included in the information asked about, without considering the content of that information (i.e., evidence or mechanism).
Cause and Effect	The information asked about includes (or is missing) some kind of causal mechanism, cause-and-effect sequence, or reason

	why things happen.
Evidence	The information asked about includes, is missing, or makes use of scientific evidence.

We are neither evaluating how the students use the ERs nor differentiating between different levels of sophistication within a particular ER. Even given free-response questions with ample writing space, middle school students often do not give enough detail in their answers to ascertain these things with any level of confidence, and this assessment was no exception. Instead, we are simply looking at whether or not a student draws on a particular ER in their answer. A difference in which ERs students draw on between sixth and eighth grades would capture a change in their framing of the questions at the most basic level, and that is what we are hoping to capture here.

Table 3 provides examples of typical responses to each question that would be assigned each code.

Table 3: Coding examples

	Code	**Example**
Question 1	Correctness	Doesn't give 100% true facts
	Amount of Information	She explains a little bit. / He explains a lot.
	Evidence	She is not showing much evidence. She didn't use proof from the data.
Question 2	Correctness	Warm water must have a lot of oxygen because I have 6 fish and they all live in 85°F water *(given as refutation of a fact in the explanation)*
	Amount of Information	This is not the best explanation because it does not have as much information as D. *(D contains the same information plus one irrelevant fact.)*
	Cause and Effect	I don't think this is a good reason because it doesn't explain why 90° isnt good.
Question 4	Correctness	The water on the plastic looks like water

	Amount of Information Cause and Effect	I would pick Makayla's model because it is more detailed then Cosette's model. She shows where it is going and how it is getting there.

Calculating causal coherence. We did not code Question 3; instead, we calculated the causal coherence of each student's explanation using the process outlined in Sandoval (2004) and Trabasso, Secco, and Van Den Broek (1984). This is a measure of how much of the explanation hangs together as a cause-and-effect chain that attempts to explain the given phenomenon. We first broke each explanation down into its constituent propositions. Table 4 shows two examples of student explanations that have been broken down into propositions. Note that in Example 2, proposition 6 does not actually help explain why the plant population would go up when carnivores are added; it is extraneous information that is simply related to the phenomenon.

Table 4: Example explanations broken into propositions

Example 1	Example 2
1. The plant population would go up 2. because the carnivores would eat the herbivores 3. which would make the herbivores population go down 4. so not a lot of organisms are eating the plants, 5. so the plant population increases.	1. The plant population will go up 2. because there are no more herbivors 3. to eat the plants, 4. which is now alowing them to grow, 5. which will bring more vegitation 6. and more oxygen will be made.

We then formed causal chains by linking together these propositions. Any propositions that are functionally equivalent were reduced to a single proposition. We then linked each proposition to any other proposition that it depends on in order to be true. Our procedure differed from that used by Trabasso, et al. in that rather than using the longest chain formed as the central causal chain, only a chain that had the phenomenon asked about in the question as its root proposition was considered to be central. Examples of propositions that would not be counted as part of the central causal chain are those that describe events that would happen after the phenomenon (e.g., #6 in example 2 of Table X), or those that say something that could be relevant to the central chain, but are not in any way connected back to it in the writing.

We then calculated the causal coherence score by dividing the number of propositions in the central chain by the total number of propositions in the explanation, giving a score from 0 to 1. This score does not reflect the correctness of an answer; it is possible for a completely incorrect answer to get a score of 1 and for an answer composed entirely of true statements including a correct mechanism for the phenomenon to get near 0. Rather than correctness, this measures how much a student's focus is on giving a causal account for the phenomenon in question as opposed to simply providing information about or related to the phenomenon.

Figure 3 shows the causal chains and coherence scores for the example explanations from Table 4. Example 1 has two propositions that are functionally equivalent: "The plant population would go up" and "so the plant population increases. Example 2 has three of these: "The plant population will go up" is equivalent to "which is now allowing them to grow" and "which will bring more vegetation." All are expressing the idea that there will be more plants. After collapsing these, each example has four distinct propositions. In this graphic, an arrow points from a proposition to another proposition that is stated to be its cause. So in Example 2, the central causal chain is made up of three propositions: There are no more herbivores (2), therefore they will not eat the plants (3), therefore more plants will grow (4). This leaves proposition 6, "and more oxygen will be made." Although the explanation connects this causally to the other propositions (more plants growing causes more oxygen to be made), the phenomenon they are meant to be explaining is the increased plant population. The fact that that increase causes something else to happen is not a part of the central causal chain explaining the increase itself. Example 1 is made up of a single causal chain and receives a score of 1, while Example 2 contains one proposition out of four (after treating propositions 1, 4, and 5 as equivalent) that does not fit into the central causal chain and thus receives a score of .75.

Figure 3: Example causal chains

Results

Overall Trends

To get an overview of the trends in the data, we combined the scores across questions for Cause and Effect, Correctness, and Amount of Information; as Evidence was only counted in Question 1, its score from that is used as its total. The one exception is the causal coherence score for Question 3; as it is not on the same scale as the others it cannot simply be added in to the Cause and Effect score, and so was included on its own. Table 5 shows the means across grades for all of these measures.

Table 5: Mean scores across grades

	6th grade	8th grade
Cause & Effect	2.82	4.25
Causal Coherence	0.75	0.88
Evidence	0.84	1.48
Correctness	2.36	1.17
Amt of Information	2.27	1.80

A MANOVA with Grade as the IV and these measures as the DVs showed an overall significant result (Pillai's Trace=.25, F=14.2, p<.001) as well as significant differences between the grades for each of the five measures, as shown in Table 6. This shows that when everything is taken into account, eighth graders talked about cause and effect and evidence more than sixth graders, and relied on correctness or amount of information less; however, the effect sizes for most of these are fairly small.

Table 6: MANOVA results

	MS	F	p	EtaSq
Cause & Effect	107.9	21.0	<.001	0.09
Causal Coherence	1.0	9.3	<.01	0.04
Evidence	22.0	43.0	<.001	0.17

Correctness	75.1	24.9	<.001	0.11
Amt of Information	11.6	3.9	<.05	0.02

Cause and Effect

Questions 2, 3, and 4 involved the practices of explanation and modeling, all of which have causal mechanism as a central concept. These would all be contexts that may activate students' *Cause and Effect* ER to answer these questions, and we expected that eighth graders would do this more often than sixth graders.

The simplest way to measure whether eighth graders called on Cause and Effect more often than sixth graders is to look at how often they mentioned it directly in their answers. Specifically, we will look at the percent of students who mentioned it *at all* (at least once) anywhere in their answers to a question, then we will look at, out of just those students who did use it at least once, the average number of times they used it across subquestions. Table 7 shows these figures for Questions 1 and 2. In both cases, more eighth graders than sixth graders used Cause and Effect at all. But of those who did use it at least once, on Question 2 both grades used it about twice with no significant difference, while on Question 4 eighth graders used it significantly more times than sixth graders.

Table 7: Frequency of Cause and Effect in Questions 2 and 4

		6th	8th	ChiSq	t
Question 2	% of students	51%	68%	8.2**	
	Mean	2.09	2.36		-1.3
Question 4	% of students	76%	90%	6.6**	
	Mean	2.4	2.9		-2.6**

*** p<.01*

Because of the way Question 2 was written, we have evidence beyond direct mentions of Cause and Effect in students' answers. The five explanations given were constructed such that some of them gave a better causal chain than others, with only B containing a full causal chain with no extraneous information. So students who are looking for causal mechanisms when evaluating a scientific explanation should be more likely to choose B as the best explanation.

Figure 4 shows the distribution of explanations chosen as the best across the two grade levels, and Table 8 shows the residual results of the Chi Squared. These two distributions are significantly different (ChiSq=14.9, p<.01). Specifically, 68% of eighth graders chose B as the best explanation, versus only 45% of sixth graders. This is largely accounted for by a significant decrease (from 21% to 6%) in students who chose E, an explanation that was basically made up of the same claim made repeatedly with different wording and no causal mechanism at all. There is also a slight decrease (from 28% to 22%) in students who chose D, the answer with extraneous information, but this difference was not significant.

Figure 4: Percent of students in each grade who chose a given explanation as the best one.

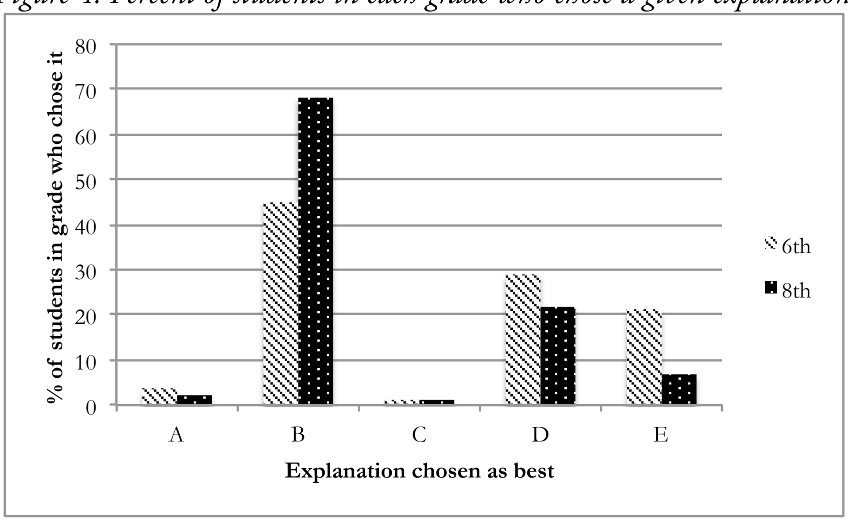


Table 8: Adjusted standardized residuals from Chi Squared of choice of explanation

	Best explanation				
Grade	A	B	C	D	E
6th	0.8	-3.6**	0.1	1.2	3.0**
8th	-0.8	3.6**	-0.1	-1.2	-3.0**

**p<.01

Though students were asked to give opinions on all five explanations, we can also look specifically at why the chose the explanation that they did as the best. When we look at what students write specifically about the answer they choose as best, we find that more eighth graders mention Cause and Effect while fewer talk about Correctness or Amount of Information (ChiSq=12.8, p<.05), as seen in Figure 5 with residuals in Table 9. This lines up with the changes in *which* explanation they chose as best, given that they moved away from choosing the ones with more correct information but not a better causal chain, towards the explanation that was long because it contained a full causal chain.

Figure 5: Reasons students gave for choosing one explanation as the best

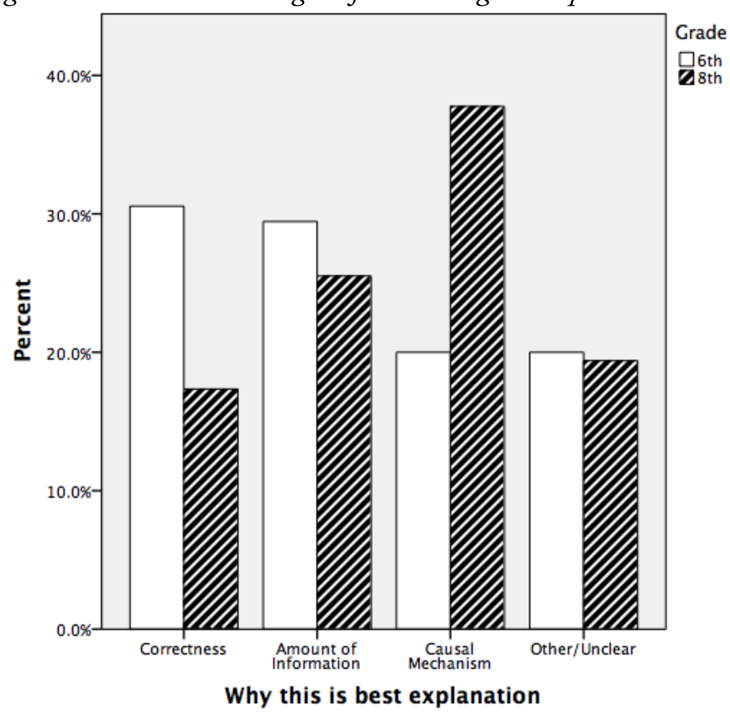


Table 9: Adjusted standardized residuals from Chi Squared of reasons given for best explanation

	Why this is the best explanation			
Grade	Correctness	Amt of Info	Cause/Effect	Other
6th	2.4*	0.7	-3.2**	0.1
8th	-2.4*	-0.7	3.2**	-0.1

**p<.05, **p<.01*

Question 3 asked students to choose one of three scientific phenomena and write a scientific explanation to explain it. We analyzed students' written explanations for *causal coherence*. This is a measure of how much of the explanation hangs together as a cause-and-effect chain that attempts to explain the given phenomenon (Sandoval, 2003; Trabasso, Secco, & Van Den Broek, 1984).

The eighth graders had an average causal coherence score of .88, while sixth graders had an average score of .73, a significant difference ($t=-3.5$, $p<.001$). This shows that the eighth graders' explanations are less likely to contain extraneous information that does not contribute to the central causal chain.

It is possible that some of the three options are easier to write a coherent explanation for, and eighth graders were more likely to choose these than sixth graders. To rule out this possibility, we ran a 2x3 ANOVA with grade and choice of phenomenon as the independent variables and the causal coherence scores as the dependent variable. The results of this can be found in Table 10. As it shows, the main effects of grade and phenomenon were significant, but there was no grade x phenomenon interaction. This indicates that while some phenomena *were* easier to write about than others, the differences between the sixth and eighth grade scores weren't significantly different across the three.

Table 10: Grade x Phenomenon choice ANOVA

	df	MS	F	p	Eta^2
Grade	1	0.80	7.86	<.01	0.03
Choice	2	1.06	10.46	<.001	0.08
Grade*Choice	2	0.21	2.04	0.13	0.02
Error	237	0.10			

Evidence

Question 1 asked students to evaluate arguments, and gave them data that they could refer to, if they chose, when doing so. One scientist (Anne) argued that new headache pills did work, but did not cite the data at all, instead referring to one person she talked to. The other scientist (Peter) cherry-picked a few points of data to argue that the pills did not work, even though the overall trend in the data is that they did. Students were asked whether they thought the pills worked and why, what the strengths and weaknesses of each argument were, and (regardless of who was right) who they thought made a better argument and why.

These questions give students a chance to use the idea of evidence in multiple ways – first in their own determination of whether the pill works, and then in their evaluations of the arguments shown. Our findings show that the picture is more complex than simply saying that eighth graders rely on evidence more than sixth graders.

When asked whether they think the pill works and why, the two groups of students are indistinguishable (see Table 11). In both grades, about two thirds of the students thought the pills worked and about one third did not (ChiSq=.823, p=.364). And in both grades, about 70% of the students relied on the raw data to make their decision, about 10% relied on the evidence presented in the arguments (Anne's anecdote or Peter's specific cherry-picked facts), and 20% relied on other information (ChiSq=.268, p=.875).

We kept this split into three categories rather than lumping the raw data and the evidence from arguments together as simply "evidence" for two reasons. First, when given the choice between data from twelve people that shows overall trends and data from one or two people, we would hope that students would base their evaluations on the former. But we also found that there were large differences between the groups of students who chose these two types of evidence; while 88% of the students who relied on the raw data came to the conclusion that the drugs worked, 90% of those who relied on the evidence in the arguments came to the conclusion that it *didn't* work (ChiSq=64.5, p<.001). So we know the choice of how to make this decision had a significant impact on what decision the students made.

Table 11: Students' assessments of whether the pills worked.

	6th	8th	ChiSq
Do the pills work?	67%	61%	0.823
Why?			0.268
Raw Data	71%	69%	
Evidence in Arguments	10%	10%	
Other	19%	22%	

These numbers suggest that most sixth graders come in understanding that this type of question should be answered based on the available evidence, aggregate data from research if possible, and that this does not change significantly for eighth graders. However, in this case, they were given little else to go on. If they had been given other types of information, such as one of the people in the argument suggesting a plausible mechanism along with their evidence, prior research suggests that they may have preferred that to the raw data (Kuhn, 2001; Sandoval and Cam, 2010). What we can really say here is that, when asked a question and presented with evidence that relates to that question, sixth graders are as likely as eighth graders to use that evidence to answer the question, are even as likely as eighth graders to choose the raw data over individual cases.

The real differences show up when sixth and eighth graders are asked to evaluate the two arguments presented. Both contain some form of evidence. Anne's conclusion (that the pill works) is correct in that it is supported by the data, while Peter's is incorrect but he provides more pieces of evidence to support it. Students were asked to list the strengths and weaknesses of each of the two arguments, then choose which argument they thought was better and explain why. The question here is whether students will evaluate the arguments based solely on their conclusion, or if they will evaluate their *use* of the evidence, i.e. the quality and/or quantity of evidence used.

We looked both at the number of times they referred to the use of evidence across all of their evaluations of the arguments, as well as specifically in their justification of whose argument was better.

Overall, eighth graders were more likely than sixth graders to mention quality or quantity of evidence at least once at any point in evaluating the arguments, and mentioned it more times total on average. 91% of eighth graders mentioned it at least once somewhere in their evaluations versus only 63% of sixth graders (ChiSq=25.1 p<.001), and those eighth graders mentioned it an average of 1.6 times overall versus an average of 1.3 times for sixth graders (t=-4.3, p<.001).

There is no significant difference in whose argument they choose as better; 91% of eighth graders and 88% of sixth graders chose Peter (ChiSq=.6, p=.74), but as shown in Figure 6, differences do arise in what students rely on specifically when supporting this choice (ChiSq=34.4, p<.001). Neither sixth nor eighth graders are very concerned with whose argument is actually correct, which is promising. While 79% of eighth graders specifically cite the quality or quantity of evidence presented, only 45% of sixth graders do. The adjusted residuals in Table 12 confirm that this is primarily offsetting a difference in the number of students who rely on the amount of information present in the argument; 29% of sixth graders rely on this while only 4% of eighth graders do. The changes in Correctness and Other are not significant.

Given that Peter does say quite a bit more than Anne total, while also presenting more evidence, it is hard to say exactly what is happening there. This could mean that the eighth graders have a better understanding of which information is important, or why (i.e. evidence is more important than general details), than the sixth graders, but it could also mean that sixth graders simply don't know how to articulate which information they care about.

Figure 6: What do students rely on to determine which argument is better?

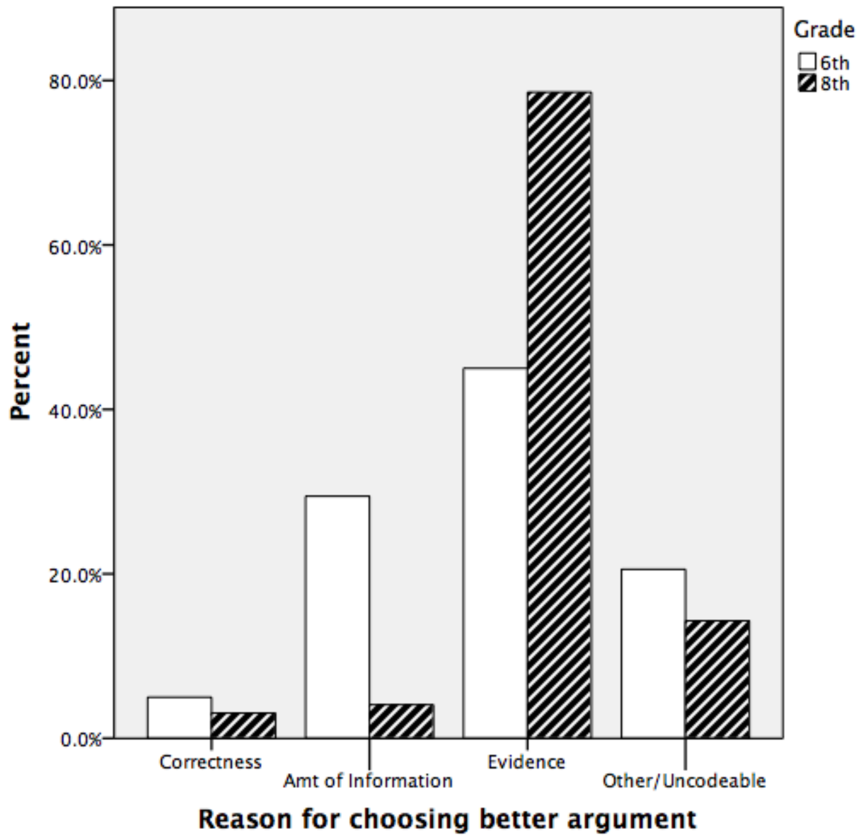


Table 12: Adjusted standardized residuals for Chi Squared of why students chose the better argument

	Reason for choosing better argument			
Grade	Correctness	Amt of Info	Evidence	Other
6th	0.8	5.0***	-5.4***	1.3
8th	-0.8	-5.0***	5.4***	-1.3

***p<.001

Correctness and Amount of Information

Questions 1, 2, and 4 found some students relying simply on whether an argument, explanation, or model was factually correct or not in their evaluations of it. Of course, there are many ways to determine whether or not something is “correct” - for example, comparing it to the evidence. However, in this case we are looking for the simplest interpretations of correctness, giving the code only when a student says something like “It sounds right” or “Because it’s wrong” without any further explanation of their reasoning.

These questions also resulted in some students judging based on the total amount of information given, regardless of what type of information that was or its structure. As with correctness, they simply stated that an argument or explanation was good because it “has enough information” or “is more detailed,” for example, without specifying what type of "information" or "details" they are talking about.

We have already seen much of these results in the previous sections about cause and effect and evidence, because often when we saw an increase in one of those it was accompanied by a decrease in these codes. In this section, we will summarize those findings again and present some new evidence that eighth graders relied on these two ERs less than sixth graders.

Table 13 shows the same analysis for Correctness and Amount of Information that we have done for the other ERs, showing the percent of students who mentioned it and of those students the mean number of times they mentioned it. For Correctness, we see significant differences in the proportion of students who used them in Questions 2 and 4, but no difference in the means. Significantly fewer eighth graders used Amount of Information on Question 1, and while the differences in the proportion of students was not significantly different in Question 2, the sixth graders used it more times per student.

Table 13: Percent of students using Correctness or Amount of Information once and mean times used

		6th	8th	ChiSq	t
Correctness					
Q1	% of students	33%	31%	0.22	
	Mean	1.1	1.1		0.74
Q2	% of students	57%	32%	16.6***	
	Mean	2.4	2.1		1
Q4	% of students	46%	15%	24.8***	
	Mean	1.3	1.1		0.98
Amt of Information					
Q1	% of students	32%	8%	20.2***	
	Mean	1.1	1		0.95
Q2	% of students	63%	74%	3.3	
	Mean	2.5	2.1		2.0*
Q4	% of students	15%	9%	1.5	
	Mean	1.3	1.4		-0.56

*p<.05 ***p<.001

Once again, we can also look at students' choice of best explanation in Question 2. In Question 2 all of the presented explanations contained only nonconflicting, factually correct information. This was intended to limit students' ability to rely on correctness as a way of evaluating the explanations, but many students still did, often exclusively. The explanations were also written specifically to differentiate between students relying on amount of information versus causal mechanism, as explanation E was as long as D and B but did not contain a causal chain (it simply restated the claim again and again in different words), and D contained extraneous information.

As we saw in Figure 5, sixth graders who chose their favorite explanation based on correctness or its having enough information were replaced by eighth graders who chose based on cause and effect, while Figure 4 showed that sixth graders who chose D or E as the best explanation were replaced by eighth graders who chose B. This picture gets even clearer when we look at which reasons students gave for choosing an explanation as best, broken down by which explanation they chose. Table 14 gives more insight into the relationship between these two figures.

Table 14 shows a breakdown of the codes that differ between grade levels, given by both grade level and the explanation chosen as best. Percentages given represent the percent of the total number of students in that grade who fall into that cell (ie, 4% of all sixth graders chose A as the best answer because it was correct). For clarity, cells which would be a 0 for both sixth and eighth graders have been left blank, and large differences (over 5%) have been bolded.

Table 14: Codes by explanation chosen as best and grade

Explanation chosen as best		Reason Given							
		Correctness		Amt of Info		Cause/Effect		Other	
	Grade	N	%	N	%	N	%	N	%
A	6	6	4%					1	1%
	8	1	1%					1	1%
B	6	19	11%	22	13%	20	**12%**	14	8%
	8	7	8%	16	18%	27	**30%**	12	13%
C	6	1	1%			1	1%		
	8	1	1%			0	0%		
D	6	4	2%	27	**16%**	11	7%	6	4%
	8	4	4%	8	**9%**	8	9%	0	0%
E	6	24	**14%**	4	2%	3	2%	4	2%
	8	3	**3%**	1	1%	1	1%	1	1%

We can see the increase in students who chose B because it mentions cause and effect, from 12% to 30%. Although B was the most popular choice for sixth graders, they were evenly split between choosing it due to correctness, amount of information, and cause and effect; eighth graders choosing B, on the other hand, clearly favored cause and effect explanations, though some still cited amount of information as well as other reasons.

This increase in students choosing B because of cause and effect appears to be offset primarily by drops in those who chose E because it is correct, which decreased from 14% to 3%, and those who chose D because of the amount of information, from 16% to 9%. These are the three largest changes, and no other changes are by more than 5% of students. Over half of sixth graders choosing D or E had given these two reasons, but among eighth graders these reasons were no higher than the others.

We saw in earlier analyses that the number of eighth graders relying on correctness or amount of information was lower than the number of sixth graders. Similarly, sixth graders choosing the answers written to appeal to students who relied on these ERs were replaced by eighth graders who chose the explanation with the most coherent causal story as the best. This breakdown shows us that, indeed these two things appear to be linked - D and E did, in fact, appeal to students who relied on correctness and amount of information, and those students were replaced by students who chose B because of its use of cause and effect.

Conclusions

In the past, many written assessments of epistemological beliefs about science have relied primarily on questions that ask the learner to directly state their beliefs. The epistemological resources framework challenges the validity of these assessments; if learners construct their epistemological framing on the fly depending on the context, making use of resources that they may not be consciously aware of, they may not be able to directly state the beliefs that underlie their behavior. If we are to analyze epistemological framing on a large scale, we need written assessments that capture behavior rather than explicit statements of belief.

This performance-based assessment captured differences in epistemological framing between sixth and eighth graders. We found that eighth graders were more likely than sixth graders to use Evidence and Cause and Effect when evaluating and creating scientific knowledge products, and less likely to use Correctness and Amount of Information. We found these differences despite the fact that we cast a wide coding net when coding Evidence and Cause and Effect, in case sixth graders were using these in simpler ways than eighth graders but still using some version of them. It is likely that looking at the same data through a narrower lens, coding only, for example, mentions of a causal mechanism specifically, would yield greater differences.

In order to accurately measure goals such as "develop[ing] students' understanding of the nature of science," we will need assessments that can be given on a large scale and that allow us to infer the beliefs that students make use of while engaging in science rather than those that they can state explicitly. This study has shown that written performance tasks such as these are useful tools for this purpose.

Design Implications

There were several limitations to this assessment design that should be addressed in future assessments. The main issue is balancing the length of the assessment with how many ERs you hope to capture and how robust you want your analysis of each to be. One constraint of using free-response items is that, particularly for younger students, they take a long time to complete. These students all had around an hour and a half to complete them (one class period at a school on block scheduling, two class periods at a school with traditional schedules), but many students were not able to finish. To differentiate between students who couldn't finish and those who did not know how to answer, we dropped any student who left more than two subquestions blank on Question 4 from the analysis of that question.

One way to create shorter assessments is to design items that afford students the opportunity to use multiple ERs. As we discovered with Question 1, though, this can get tricky if what you want students to notice is the lack of something. In this case, we'd hoped that some students would note that neither person gave any causal mechanism as part of their argument. Only two eighth graders mentioned this at all. This could mean that even the eighth graders were unaware that one could evaluate arguments based on their inclusion of a causal mechanism.

But there were many students who evaluated the explanations in Question 2 based on their inclusion of a causal mechanism yet did not mention it in Question 1. This implies that the problem may be in the design of the questions rather than the students' understanding of scientific argumentation. In Question 2, students are presented with alternatives that vary according to specific traits, including causal mechanism. When looking at two of these explanations side-by-side, it is easier to notice that one is lacking something the other is not. In Question 1, neither person includes causal mechanism, but one person makes use of the given data and one does not. Students' attention, then, is drawn to the differences between the two arguments rather than something that is not there at all.

So which is better-designed? That depends on your assessment goals. For our purposes, trying to cast a wide net and catch as many students using the more productive ERs as possible, looking simply to demonstrate that students *can* frame these activities productively and that we can capture that, Question 2 is more useful. For this study, if a student has an epistemological frame to draw on that includes the use of these ERs in the appropriate context, we want to activate that frame. And younger students may have less stable frames, and may need to have each ER activated separately. The comparison to a more complete explanation highlights the other explanation's lack of causal mechanism, giving an extra prompt for students to activate an epistemological frame involving evaluation explanations based on causal mechanism. A student who does not have such a frame will not see any meaningful difference between the two explanations.

Question 1, on the other hand, may be more appropriate for more advanced students whose epistemological frames are more robust and stable. While providing the comparison gives as many students as possible the opportunity to activate the appropriate ER, not highlighting the lack ensures that only those students whose epistemological framing of the situation is robust enough to include those ERs that are not explicitly prompted along with those that are will notice it.

Another concern competing with length is the desire to have multiple data points about each ER. We have far more data about Cause and Effect than about Evidence; three questions' worth vs one. Within Question 1 we do have multiple subquestions, so students may mention Evidence in multiple places and multiple ways, but we did not design opportunities to activate Evidence into any of the other questions. Given that our assessment was too long for some students to complete, in retrospect it seems like a waste to design Question 4, which was a long item in itself, to only directly prompt ERs that were also prompted by Question 2. On the other hand, because students often gave very short answers that may not reflect the entirety of their thinking, it is useful to give them multiple chances to show that they used a particular ER.

In the end, the balance of length versus robustness depends on the goals and context of the assessment. Perhaps Question 4 could have been rewritten to include evidence, or perhaps it could have been shortened significantly because we already had data on Cause and Effect and Correctness. We could rewrite Question 1 in the future to include a causal mechanism in one of the arguments, but this would make the item longer and more complicated. It may have become too difficult for sixth graders to read and understand, confounding the data.

In the future, it would be useful to be able to analyze how students make use of these ERs instead of simply which ones they use, as it is possible for two students to use the same ER in very different ways, one of which may be much more productive than the other (Berland & Crucet, 2015). Unfortunately, this is difficult with the current data, because students mostly gave very short answers, providing minimal detail even when prompted to explain their reasoning. This is a common problem with written assessments, and researchers often compensate by pairing written assessments with clinical interviews that give students a chance to expand on their reasoning. While items that successfully prompt students to provide sufficient detail will result in a longer assessment, they may also provide more opportunities for students to use a particular ER, giving more data both about how likely it is to be used and how it is used.

Chapter 3:

The Impact of Journalistic Choices on Students' Reasoning About A Socioscientific Issue in News Articles

In his 1916 speech to the National Education Association, Dewey stated that the goal of science education should be scientific thinking, and that if a man isn't familiar with "scientific ways of dealing with a subject matter, he has no sure way of telling the difference between all wool knowledge and shoddy goods. (pp 3-4)”

Since then, a variety of models of scientific literacy have emerged (Shen, 1975; AAAS, 1985; Miller, 1998; NRC, 2006; 2009; OECD, 2006), all of which include the idea that not only scientists, but any scientifically literate citizen should have some understanding of how science works and that this knowledge is useful in people's day-to-day lives. The Next Generation Science Standards state that understanding science is key to "comprehending current events, choosing and using technology, or making informed decisions about one’s health care. (NGSS Lead States, 2013, 3)" As scientific issues such as evolution and sustainability become cultural debates that every member of our society must grapple with, it is more important than ever that we prepare students to face these socioscientific issues (SSIs) as adults.

Most of them will not be participating in these debates as scientists. They will neither have access to, nor the specialized knowledge required to read, the original journal articles where the research in published. Instead, they will mainly rely on media reports for their information about current events related to science. In 2014, 28% of Americans used the television as their primary source of science information, 21% news sites on the internet, and 13% newspapers or magazines (National Science Board, 2016). 15% cite a search engine such as Google, rather than the eventual source of their information; likely some of that 15% wind up on a media site as well. It is so common for adults to get their scientific information primarily from the media that some suggest we choose our topics for assessment based on what is covered in the news (Tseng, Chang, Rundgren, & Rundgren, 2010).

Getting information from the news media is not the same as getting it from a textbook or journal article. As Maier, et al. (2014) make clear, the process of informal science learning through news media involves not only the reader's selection of sources and processing of the content, but also the journalist's selection and depiction of the science topics. The journalist's beliefs about their own role, their ideas (and their editors' ideas) about what is newsworthy, and journalistic norms such as balance or particular framing devices all play a role in how science is depicted and thus what people learn when they seek scientific information from the news media.

In this study, we will explore how students bring their understanding of science to bear when reading about a SSI in the news media and how journalistic choices affect that process.

Theoretical Background

Socioscientific Issues

Socioscientific issues are "societal dilemmas with conceptual, procedural, or technological links to science" (Sadler & Zeidler, 2004a, p. 5). While all science is inextricably linked to the society that produces it, these are issues that "display a unique degree of societal interest, effect, and consequent" (p. 5). Some examples that have been studied in the science education literature include genetic engineering (Sadler & Zeidler, 2004a); the health impacts of various types of radiation (Kolstø, 2001; Albe 2008a, 2008b), the spread of disease (Lee & Grace, 2012), water quality (Sadler, Barab, & Scott, 2007), global warming (Sadler & Zeidler, 2004b), and more.

When making decisions about SSIs, people do not attend solely to the scientific evidence, both for systematic reasons and due to lack of understanding. Sadler, Chambers, and Zeidler (2004) found that only about half of the high school students they studied showed a complex understanding of was scientific data was and how to use it in their argument, and 17% seemed confused about what counted as scientific data at all. Sadler & Zeidler (2004a), investigating the moral dimension of SSIs, found that college students relied on consequential moral reasoning, moral principles, emotional reactions, and moral intuition when talking about genetic therapy and cloning. Albe (2008b) found that French high school students engaging in group discussions drew on "common knowledge," or prior knowledge that they introduce to the discussion with the assumption that the others will also know it and take it as true. Kolstø (2001a) looked at how Norwegian high school students evaluated the trustworthiness of knowledge claims about a SSI and found that they were more likely to evaluate the sources of the information than the content of it; if the source seemed reliable, they would accept the information with no further proof needed.

Tytler and Duggan (2001), in a case study of three people directly involved in a dispute about burning recycled liquid fuel in the UK, along with an analysis of related public documents, found that three different categories of information were used as evidence in the public debate. In addition to scientific evidence based on actual data, people also presented informal evidence such as personal experience, common sense, or circumstantial evidence. They also presented as evidence information intended to frame the issue, rather than to support or dispute the scientific evidence, such as the effects on jobs in the area or the legal rights of the different groups involved.

These studies all point to the significant differences between an SSI and a normal topic of scientific investigation. Both in classrooms and in real life, people do not make decisions based solely on the scientific evidence, as scientists would when investigating a phenomenon. These issues are complex and tied to many segments of our society, and people must consider the impact on various people and groups not only scientifically but economically and morally as well. But it also shows that people sometimes give information that is less reliable or generalizable, such as individual experiences or common sense, equal weight to scientific data when they may not be comparable.

In addition to looking at how people reason about SSIs, researchers have looked at the factors that impact that reasoning. Lee & Grace (2012) studied the impact of cultural differences between students in Hong Kong, which was under British rule for many years, versus Guangzhou, a nearby city on the mainland, having them undergo a two week activity exploring and discussing ways to prevent the spread of avian flu. They found that the cultural differences between the two groups affected the ways they framed the issue both before and after the activity, how much the activity changed those frames, and what information they deemed necessary to address the issue.

Some studies have found that epistemological beliefs, either in general or about science, impact reasoning about SSIs, while others have struggled to make this link. Schommer-Aikens and Hutter (2002) found that the more people expressed a belief that knowledge is tentative and complex, the more likely they were to be able to take on multiple perspectives about a controversial scientific issue, to withhold their decision until all information was available, to modify their thinking, and to acknowledge the complexity of the issue. Sadler, Chambers, and Zeidler (2004) found that students brought a variety of beliefs about the nature of science (NOS) to bear in their reasoning about SSIs, including their understanding of scientific data, the social embeddedness of science, and the tentativeness of knowledge.

But when studies have looked for a connection between explicit assessments of views of science and SSIs, they have largely fallen short. Liu, Lin, and Tsai (2011) assessed five dimensions of science epistemology and tried to correlate them to four modes of thinking students used when reasoning about an invasive species. Out of twenty possible dimension x mode of thinking relationships, the only significant relationship they were able to find was that those students who professed a stronger belief in the "invented and creative nature of science" were less likely to reason from an ecological perspective, a relationship for which the authors did not have a solid theoretical explanation. Zeidler, Walker, Ackett, and Simmons (2002) provided detailed taxonomic analyses of student's conceptions of NOS and their responses to SSIs, but were able to find "only a few discernible instances of a clear relationship" (p. 359) between the two. Zeidler, et al (2013) found cultural differences between five countries in terms of how students reasoned about an SSI and their NOS beliefs, but although they tentatively infer indirect support for a relationship between the two they do not do the analysis needed to actually determine if this is true.

One thing many studies of reasoning about SSIs have in common is that they give their participants constructed texts to learn about and reason about the SSI, rather than real media reports. This is understandable, as researchers would want to have tight control over the content provided to participants, in terms of what sort of evidence is used, what perspectives are included, etc. But given that most people will get their information about SSIs from the media in real life, it ignores many of the complexities that set science journalism apart from other science texts, from the purpose of the articles to their structure and how the issues are framed.

Science Journalism

Media reports of science can be quite different from academic science texts, even those reporting on a single study rather than a controversial issue. Jiménez-Alexandre and Frederico-Agraso (2009) analyzed the structure of a biology article published in Nature and three journalistic reported versions (JRVs) of the article. They found that while the Nature article made one claim that it backed up with evidence (the study the article was about) and a secondary, more general and sensational claim that it did not provide any evidence for, three of the four JRVs used the secondary, unsupported claim as their headline. All four mentioned the main claim briefly, but tended to omit nearly all of the evidence in support of it; they included diagrams or micrographs of cloned embryos from the article, but not the results of the experiments they were from. In the rare cases when some of the evidence was included, it was used to support the secondary claim even though it was not used in that way in the original article. They all included the original article's justifications in support of its secondary claim. Overall, all four articles both devoted more paragraphs to the secondary, unsupported, sensationalistic claim and structured their arguments around it.

In journalism studies, it has long been accepted that the mass media has a major impact on what information the public has access to and how they think about that information. White (1950) took this view of journalists and others in media as "gatekeepers" of information, and showed how one newspaper editor's personal tastes, opinions, and biases influenced the content of his newspaper, and thus the information that people in his small midwestern city had access to. He chose which of thousands of column inches of copy that came over his wire made it into his paper, accepting only about 1% of the total, and his personal opinions determined not only which stories received more attention, but which perspectives on those stories were shown and how the topics were written about.

Not only are media outlets seen as gatekeepers of information, they are also theorized to play a major role in agenda-setting, in that the issues that the media chooses to emphasize tend to be seen as more important by mass audiences (Scheufele & Tewksbury, 2007). Although it is true that media likely choose to cover issues that their audience wants to hear about, it is also theorized that repeated exposure to particular issues via the media makes those issues more salient and accessible in people's minds. In particular, some media outlets are considered to be "elite" or "prestige" due to having a large, nationwide audience and high levels of perceived legitimacy (Deephouse & Suchman, 2008), and may be particularly instrumental in agenda-setting on a national level (Boykoff & Boykoff, 2004).

All of this means that media accounts of science can have a major impact on society's views of science, and that the personal biases of individual editors and journalists, along with wider news media norms and trends, will shape the information the public has about SSIs and scientific controversies and how these issues are framed.

Stocking and Holstein (2008), for example, looked at one case where there was little doubt as to the results of the scientific research, but it was virulently opposed by an industry that stood to lose profits because of it, and how the local media presented this information to the public. Through analysis of media and industry documents and interviews with individual journalists, they found that the journalists' perceptions of their role as a journalist largely determined how much attention their stories gave to the industry's denial of the scientific evidence. Some saw themselves in roles that required they present all available information to the public and let the public draw their own conclusions, and so they printed the industry's critical comments. Others saw it as part of their responsibility to evaluate the information for themselves and only present that which they found to be valid and WORD, and so when they found the industry's critiques of the research to be baseless, they did not print them. Whether the public was presented only with valid scientific evidence, or also with industry's attempts to discredit those results depended entirely on what news they read and how that particular journalist construed his or her own responsibilities as a journalist.

On a larger scale, science journalism tends to follow the same journalistic norms as any other topic. Journalists are writing to get the attention of their audience, to convince people to read their story and not the story in the next newspaper or website, and so they rely on traditionally attention-grabbing themes such as timeliness, conflict, and novelty (Dunwoody, 2014). These have a clear effect on what science topics are written about; for example, journalists almost always cover the end product (a study that is complete and about to be published) rather than showing science in progress. The attention-drawing potential of conflict, especially, can lead journalists to, intentionally or not, create the appearance of scientific debate where there is none—particularly when it is combined with another journalistic norm, that of balance (e.g. Dearing, 1995; Stocking & Holstein, 2008; Clarke, 2008).

Journalists strive for balance in their articles, in that they want to present all sides of a controversy, giving them equal weight rather than being biased toward one. When all of these sides are political positions, or personal opinions or experiences, this can work well to be sure that they are giving an accurate picture of the debate. However, when covering an issue where the scientific evidence is all on one side, this often results in journalists giving equal weight or coverage to fringe positions with little to no scientific evidence to support them, simply because there is no other way for them to "balance" their coverage with multiple perspectives. Researchers have found that this attempt at balance results in coverage that does not reflect the scientific consensus on topics as wide-ranging as global warming (Boykoff & Boykoff, 2004), hog farming (Stocking & Holstein, 2008), the cause of AIDS (Dearing, 1995), cold fusion (Dearing, 1995), and whether vaccines cause autism (Clarke, 2008).

Given the earlier points about the news media acting as gatekeepers of information and setters of the public agenda, this can easily result in the public perceiving there to be legitimate scientific debate where none exists. Dixon and Clarke (2012) found that people who read a "balanced" article (one that mentioned evidence or expert claims both in favor of and against the vaccine-autism link) were less certain that there is no link, less certain that scientists believe there is no link, and more likely to think there is a divide in the medical community on the issue than those who read articles that only included anti-link evidence. In fact, those who read the balanced article were no different from those who read an article that included only pro-link evidence on these measures.

This effect may be exacerbated by the fact that journalists tend to rarely address the uncertainty of scientific research unless they are specifically framing their story to highlight it, as in the case of a controversy or debate (Maier, et al, 2014; Nisbet & Scheufele, 2009). This could give the impression that such uncertainty is a rare occurrence that implies a legitimate debate, rather than an inherent part of the scientific process.

Uncertainty is one of several frames commonly used by science journalists to set the tone and perspective of their stories. In media research, the term framing is used to describe the process of organizing a story according to particular themes; a frame "stresses certain aspects of reality and pushes others into the background" (Ruhrmann, Guenther, Kessler, & Milde, 2015, p. 684), determining what information and perspectives are included and emphasized and what is left out. Nisbet & Scheufele (2009) point out that over time, researchers have converged on a common set of frames that tend to reappear regularly in stories about science-related policy debates: social progress, economic development/competitiveness, morality/ethics, scientific/technical uncertainty, Pandora's box/Frankenstein's monster/runaway science, public accountability/governance, middle way/alternative path, and conflict/strategy.

Ruhrmann, et al (2015), rather than looking at these overarching frames in television reports of science, focused on how they framed scientific uncertainty. They found four common frames: controversy, scientifically certain data, everyday medical risks, and conflicting scientific evidence. These varied in how much they mentioned uncertainty, to what the uncertainty was attributed, who was shown or spoke onscreen (scientists, doctors, patients, etc), what other themes were emphasized (such as risks, benefits, or social problems), and whether the overall tone was positive or negative.

Journalism also differs from general science writing in the structure of the articles. Many news articles use an "inverted pyramid" structure where information is presented in order from most to least important rather than organizing information conceptually or chronologically. This is so that readers are drawn in by the most interesting, sensational, or useful information, ensuring an audience. Yaros (2011) took science news articles written in this format and rearranged them into a more traditional linear narrative format. He found that while participants rated the inverted pyramid articles as more interesting, they understood the linearly-written articles better. This implies that even when science news is framed in a way that accurately represents the science, readers may not understand the content as well as they would coming from another source.

Halkia and Mantzouridis (2005) also looked at how journalistic choices affect interest. They found that Greek secondary students would rather read articles with more sensational headlines that focus on some danger to humanity over those with less exciting headlines. When asked why they chose the article they did, they cited the danger aspect or the headline more generally more often than general interest in the scientific subject of the article. After reading the article they chose, when asked what they liked most about it, they cited the didactic tools used (such as analogies and metaphors) and the illustrations the most, followed closely by statistics about some kind of danger, and were less likely to cite the views and activities of scientists written about, technological issues, or scientific models/theories.

Going back to their study of JRVs of a Nature article, all of which focused on the article's sensational yet unsupported claim over its less attention-grabbing but better-supported claim, Jiménez-Alexandre and Frederico-Agraso (2009) found that when college students were asked to summarize one of the JRVs, they tended to include both claims as well as the evidence and justifications used in the JRV, which were mostly used to support the unsupported (in the original article) claim. But when asked to provide reasons for or against this type of research, they overwhelmingly cited its potential use in therapy (the unsupported, sensational claim), implying that the JRV's emphasis on the sensational aspects of the original article had an impact on their thinking about the research.

All of this suggests that, if the purpose of a news article is fundamentally to sell newspapers or advertising space on a web page, the structural and linguistic tricks that journalists use to pique readers' interest do their job, even if that might come at the expense of readers' understanding the science content they are writing about.

Epistemological Resources

This research views beliefs about what scientific knowledge is and how it is constructed through an epistemological resources (ER) lens. Unlike much of the work on socioscientific issues, which treats beliefs about science as explicit statements that can be verbalized and are stable across contexts (e.g. Liu, Lin, and Tsai, 2011; Zeidler, et al, 2013; Leung, Wong, & Yung, 2014), the epistemological resources framework conceptualizes these beliefs as implicit, making them sometimes difficult to verbalize, and highly context-dependent (Hammer and Elby, 2002; Hammer et al, 2004).

This conceptualization helps to account for results like those of Salter and Aikens (2013), who found that while students' results on an NOS assessment did not change over the course of a semester, their performance during classroom activities showed an increase in sophistication. Bell and Lederman (2002) also found that a summer program working in science laboratories did not change students' explicitly-stated beliefs, despite the fact that their mentors expected to see growth based on their performance in the labs. These students may be using more sophisticated combinations of epistemological resources to engage in classroom or laboratory activities, even when their ability to verbalize explicit beliefs has not yet changed, or they may view the assessment and the activity as sufficiently different contexts that they do not activate the same resources.

Similarly, this could account for the difficulty some researchers have faced in linking NOS assessment results with students' performance when reasoning about, discussing, or debating SSIs (Zeidler, Walker, Ackett, and Simmons, 2002; Liu, Lin, and Tsai, 2011; Leung, Wong, & Yung, 2014). The assessments that rely on explicit statements of belief are a very different context from activities around SSIs, and may not prompt students to activate the same sets of resources. This is supported by the fact that researchers have had more success when looking for evidence of beliefs in students' reasoning rather than correlating it with a separate assessment (Sadler, Chambers, and Zeidler, 2004).

If we want to know whether the understanding of science we are building in the classroom will help students when they encounter science in the media, looking at explicit statements of beliefs is likely the wrong approach. An ER perspective suggests that we need to assess within the context we care about. As we've seen, news articles about science can differ greatly from other science writing, so if this is the context we are curious about we need to use actual news articles. And when drawing conclusions about students' thinking, we need to pay attention to ideas that students may not explicitly verbalize.

Research Questions

Given this background, this study explores the following questions:

When reading a news article about an SSI, what do middle school students find most convincing?

How do they make use of their understanding of scientific evidence when reasoning about the SSI in the article?

How do the journalistic features of the articles affect students' thinking?

Methods

Participants

Participants were 152 sixth graders, 69 seventh graders, and 83 eighth graders at two midwestern middle schools. Problem with consent forms led to fewer seventh and eighth graders being included than sixth graders.

Assessment Design

Choice of articles. This study was designed in the spring of 2010; Andrew Wakefield's study purportedly linking the MMR vaccine to autism had only been deemed fraudulent and retracted by The Lancet that January (The Editors of the Lancet, 2010), and this potential connection was still very much a topic of debate in the news media, if not in the scientific community. Clarke (2008) had already shown that, probably in an attempt to provide "balanced" coverage, news articles contained far more evidence and expert claims supporting the vaccine-autism link than the scientific consensus would suggest. In 2012, Dixon and Clarke went on to show that this "balanced" reporting reduced readers' certainty that vaccines do not cause autism as much as reporting that showed only evidence in favor of the link. This topic was chosen as a SSI that was timely, and likely to be misrepresented in the news media.

When looking for articles about this topic, we looked at articles that were easily available online on news sites when we performed a Google search for "vaccines and autism," making them ones that students would be likely to find if they were researching the topic. We also looked for two articles that represented very different ways of presenting the debate, while both still containing at least one expert supporting each side. Clarke (2008) only looked at whether or not articles included each side, not how much space each side was given or how they were presented. We wanted to see if, given two "balanced" articles, other journalistic choices might influence students' thinking.

The two articles we selected will be referred to as Article A, found on CNN.com (Martin, 2008), and Article B, found on Bloomberg.com (Cortez, 2006). These articles differ in how much space they devote to the two sides of the debate, how they frame the issue, and the writing styles they use, as well as what types of information they present in support of the pro-link side of the debate (in both cases, scientific evidence is provided in support of the anti-link side). Article A, in general, gives more support to the pro-link side of the debate than Article B.

Article A's headline, "Vaccine-autism question divides parents, scientists" does not endorse one side of the debate or the other. It also does not make it clear who is divided—while the article indicates that parents are divided from scientists, the headline can be interpreted as indicating that both groups, parents and scientists, are internally divided. Both of these features of the headline imply more controversy than actually exists among scientists. Article B's headline, on the other hand, "Autism's Link to Vaccines Is Unfounded, Study Says," focuses entirely on the anti-link position.

Article A devotes fourteen out of thirty-six sentences, or 273 out of 776 words, to the anti-link side of the debate. That's 39% of its sentences and 35% of its words in support of the side that agrees with scientific consensus. Article B, on the other hand, Article B gives the anti-link side eleven out of nineteen sentences (58%), or 284 out of 479 words (59%). Article A is clearly biased in terms of how much space it devotes to the pro-link side; Article B is more balanced, only slightly preferring the anti-link side that its headline supports. This means that Article B still gives much more attention to the pro-link side than is supported by scientific evidence.

Ruhrmann et al's (2015) study of the presentation of scientific uncertainty gives us a useful way to look at how these two articles frame the issue. Article A fits into their Everyday Medical Risks frame; it allows medical doctors and patients to speak more than scientists, it relies on personal experiences, it focuses on the risk of vaccines causing autism over the benefits of vaccines, and is negative toward the science in its overall tone. Article B fits into their Conflicting Scientific Evidence frame; it mainly quotes scientists or representatives of scientific institutions and its tenor is more balanced. Though while it does focus on scientific uncertainty, this uncertainty is shown to arise not from conflicting results but from conflicting interpretations of the results of one study, which is slightly different from the norm for this frame.

The articles also show evidence of their authors perceiving themselves as falling into different roles as used by Stocking and Holstein (2008). Article A's author appears to see himself as a Populist Mobilizer; these journalists "engage audiences with entertaining stories that give a voice to the public and set political agendas (p. 10)." Like the Populist Mobilizers in Stocking and Holstein's study, the author of Article A focuses on the perspectives of ordinary people directly impacted by the issue. Stocking and Holstein's Populist Mobilizers saw it as their duty to present the argument against the science so that people could draw their own conclusions, though since the science supported the views of the victims this often led to more weight being given to the science than the industrial opposition. In this case, the victims' views are in direct conflict with the scientific evidence, so it is difficult to draw direct parallels, but the author of Article A did make use of the science expert that was offering support for the victims' claims.

Article B's author, on the other hand, seems to view herself more as a Disseminator, whose job is "getting the straight facts and getting them to the public quickly (p. 10)." The Disseminator seeks to publish the views of all parties, even if one side is more credible, trusting the public to make that determination for themselves. Article B follows this pattern, reporting on the results of the study but also giving voice to an opponent who claims that the author of the study is biased and that the results can be interpreted differently. Though Article B gives slightly more space to the anti-link side supported by the study, it does not explicitly judge which side is more credible.

One last difference is their reliance on narrative versus expository writing. Though Article A opens with the most eye-catching and important facts, as the Inverted Pyramid would suggest, it then segues into a narrative, devoting three paragraphs to the story of one autistic girl before returning to an expository style for the rest of the article. Article B, on the other hand, is entirely expository. Moyer-Gusé and Nabi (2010) found that the use of narrative over expository style can reduce resistance to persuasive content. While Article A still contains much more expository than narrative prose, its use of narrative on one side of the debate and not the other could play some part in swaying the reader.

The two articles were edited only lightly, to clarify some terminology and to cut out some information that was irrelevant to the scientific debate itself (such as some information on the specifics of legal proceedings).

Assessments utilizing the two articles were distributed randomly in each class. 72 sixth graders, 31 seventh graders, and 48 eighth graders were given Article A, while 80 sixth graders, 38 seventh graders, and 35 eighth graders were given Article B. The assessment with Article A can be found in Appendix B, and the assessment with Article B can be found in Appendix C.

Questions and tasks. After reading the article, the students were asked the following questions:

1. After reading this article, do you think that vaccines cause autism? Yes / No Why or why not?

2. Did anything you read in the article affect your answer to Question 1? If so, what?

3. Do you think that there is more scientific evidence that vaccines do cause autism, or that they do not? Do / Do Not Why?

They were given ample space for their free-response answers, and told to use the back of the page if needed.

After answering these questions, they were instructed, "Please underline any parts of the article that you think give scientific evidence in either direction. (Scientific evidence that vaccines either do cause autism, or that they don't.)"

This underlining task was placed after the questions so that they would not first underline scientific evidence, and then have the parts that they underlined more readily come to mind when trying to answer the questions. It is also important to note that we did not ask them to underline the parts that they found convincing, or those that they used in answering the questions; in fact, we asked them to underline scientific evidence no matter what side of the debate it supported.

In most studies that analyze students' reasoning about SSIs, the researchers are the ones who classify the information students use in making their decisions as scientific evidence versus other types of evidence or information (Sadler, Chambers, and Zeidler, 2004; Tytler, Duggan, and Gott, 2001). When we ask how students are using their understanding of scientific evidence, we don't want to assume what that understanding is. We do want to know whether they are using what we would classify as scientific evidence, but we also want to know how they understand themselves to be approaching this task—do they believe themselves to be basing their decisions on scientific evidence?

Asking them directly would pose many problems—they may not be able to articulate a definition of scientific evidence, or they might, upon being asked about whether their answer was based on scientific evidence, realize that the "right answer" is yes and try to come up with a post hoc justification for that. Instead, we are asking them to justify their answer, then asking them separately about where they think the scientific evidence points, and then asking them to classify the content of the article as scientific evidence or not, without asking them to revisit or justify their original answer in light of their thoughts on scientific evidence. We will then look at the relationship between how they answer these questions and how they classify the content of the article to determine whether they were using what they believe to be scientific evidence in their reasoning.

The assessment included several more questions and another classification task, but these will not be analyzed in this paper.

Coding

It was always our intention to collapse questions 1 and 2 for coding, as question 2 is really just a way to prompt students to elaborate further on their reasoning for their answer in question 1. In the end, we wound up collapsing all three questions.

When we began coding, the first thing we discovered was that about a third of students completely misunderstood question 3, "Do you think that there is more scientific evidence that vaccines do cause autism, or that they do not?" We had worded this carefully to avoid confusion, and the students we piloted the assessment on had no problem with it. But a third of students clearly did not understand the question to be asking them "for which side have scientists found the most evidence so far," but "do you believe that more evidence is out there to be found?" This resulted in answers that were not relevant for our purposes, and could not be compared to the answers that correctly understood the question. Losing a third of the data resulted in few enough answers that it would be difficult to compare across grades or between those who answered "yes" versus "no" to the original question. Many of the remaining answers were not detailed enough to code or simply restated something from the first two questions. In the end, we wound up collapsing question 3 in with 1 and 2.

We coded the assessments of Article A first. We began the iterative process of establishing codes with some ideas of the types of concepts we expected to see students drawing on in their responses such as the anecdote given in the story or the scientific evidence cited in the story, but our initial pass through the data attempted to note any and all ideas used in students' answers. From there, similar codes were grouped together into more general codes—for example, some students talked about the idea that mercury in the vaccines causes autism, while others talked about the measles virus itself causing autism, and both of these were collapsed into the category "Mechanism" along with other instances where students talked about how, specifically, the vaccine (or something other than the vaccine) might cause autism.

We continued this process until we were satisfied that every substantive answer had been assigned at least one code (many were given more than one, as the students talked about multiple ideas), there was no redundancy in the codes, and every code was one that applied to at least several answers (the rarest was used nine times).

There was one time in this process when, instead of combining similar codes, we had to separate a code into two codes because we realized that, while the codes involved students citing the same fact (that very, very few children who receive an MMR vaccine go on to develop autism), they were interpreting that fact in two entirely opposite ways using different logic. This resulted in the codes Lack of Correlation, referring to students who claim that the vaccine must not cause autism because so few children who get it develop autism, and Rarely Causes, for students who instead conclude that the vaccine does cause autism, but only on rare occasions.

The end result of this process was a set of nine codes:

Anecdote - The answer cites one of the anecdotal stories in the article.

Research - The answer cites the scientific research mentioned in the article.

Appeal to Authority - The answer cites something that a doctor, scientist, or other expert said, or what the experts "say" in general, but not the research behind that claim.

Correlation - The answer mentions a positive correlation between the MMR vaccine and autism, such as the fact that some children develop autism immediately after receiving the vaccine.

Coincidence - The answer writes off any correlation between the two as a coincidence.

Mechanism - The answer discusses a possible mechanism explaining how the vaccine or another source could cause autism.

Lack of Evidence - The answer claims that one side of the argument must be correct because the other side does not have enough evidence, without talking about the evidence supporting the side that is supposedly correct.

Lack of Correlation - The answer points out that the vast majority of children who receive an MMR vaccine do not develop autism, therefore the vaccine could not be the cause.

Rarely Causes - The answer claims that the MMR vaccine causes autism, but only every once in a while or under very specific, rare conditions.

When we began the process of coding Article B, we started out with this list of codes, knowing that some would not apply (there was no anecdote in the article, for example) and that new codes might be needed.

We kept the Research and Mechanism codes, and added three new ones:

Personal Experience - The answer draws on the student's own experiences with vaccines or autism.

More Evidence/Information - The answer mentions one side having more "evidence" or more "information," but does not make it clear exactly what they are referring to. Despite the use of the word "evidence" in some of these answers, we did not assume that this was referring to the scientific research unless it mentioned the research specifically; as one of our questions is how students are thinking about evidence, we did not want to assume that they defined this word the same way that we would.

Discrediting - The answer refers to one scientist's attempts to discredit the other scientist and his work.

We also noted that many students were misinterpreting the research and statistics cited in the article; for example, one student wrote, "I think vaccines cause autism because in the time period that people took the vaccines, autism rates increased. … The part that stated that there was an increase of people born between 1987-1998 who had autism convinced me that the shots caused it. The increase was evidence or data in which supported that autism was caused by these shots." This person correctly notes that autism rates rose during the study, but either did not notice or did not understand that mercury was removed from vaccines midway through this time period, yet rates continued to rise, indicating a lack of correlation. We added a code for this, though since it is not a distinct category of content it will be analyzed separately.

Data Reduction

Article A. There are thirty-six sentences in the article, and the initial coding pass involved twelve codes. Given that the data set is only 154 students (only 114 of whom underlined any sentences at all), many of these were underlined or used by too few students to be useful in an analysis. Using this many non-mutually-exclusive categories, because significance must be tested separately for every interaction among them (ie sentence 1 x research, sentence 1 x anecdote, etc), would also necessitate so many chi squared tests that even at a .01 significance level we would expect several false positives.

To reduce the data, we looked at two ways we were planning to break down the data in analysis: by grade level, and by students who do or do not believe that vaccines or measles cause autism. We kept codes and sentences that were underlined or used by at least 25% of the students in at least one grade level, and 25% of the students who either do or don't believe, to ensure that every analysis we run has sufficient n in each cell to be meaningful.

This system left us with six codes: Research, Coincidence, Anecdote, Correlation, Mechanism, and Authority. It also left us with eleven sentences: 7, 8, 9, 10, 14, 15, 20, 23, 28, 29, and 30.

These sentences tend to fall into groups in the text, often several within one paragraph. It is not always interesting which of these specifically a student underlines. For example, if one student underlines sentence 9, "Within days of receiving the injection, Michelle suffered from a high fever, persistent vomiting and problems with her digestion," while another student underlines sentence 10, "Worse still, her parents say, Michelle stopped speaking and no longer responded to her name," both of which are pieces of anecdotal evidence supporting the claim in sentence 8, there is not necessarily a meaningful difference between the two, or between them and a third student who underlined both. The sentences are both functionally the same and referring to the same topic within the article.

For some parts of the analysis, then, we have further grouped these sentences into episodes. Sentences 7, 8, 9, and 10 are "Anecdote 1", 14 and 15 are "Science 1", 20 and 23 are "Science 2", and 28, 29, and 30 are "Anecdote 2". This was initially based on their placement, all groups being sentences that are in the same paragraph or that refer to each other in some way, but factor analysis with varimax rotation confirmed that these groupings share factor loadings, and these loadings are distinct among the four groups. Table 15 shows these factor loadings (the factor analysis was done using all 36 sentences, but only the eleven underlined by enough students for this analysis are shown in the table). As the table shows, most of the lines in each anecdotal episode share a secondary loading with the other anecdotal episode's primary loading, and the science episodes follow the same pattern, evidence that these four groups are made up of two sets of similar groups.

Table 15: Factor analysis showing sentences that share factor loadings.

	Factor			
Line Number	1	2	3	4
7	0.035	0.133	0.442*	0.047
8	0.106	0.262	0.45*	-0.052
9	0.061	0.123	0.792*	-0.079
10	0.09	0.085	0.839*	-0.063
14	0.579*	0.031	0.186	0.214
15	0.717*	0.032	0.015	0.197
20	0.184	0.011	-0.076	0.534*
23	0.067	0.083	-0.257	0.549*
28	0.164	0.609*	0.302	0.223

29	0.01	0.552*	0.245	-0.012
30	0.03	0.638*	0.364	-0.038

** This line's highest factor loading*

Our analysis, therefore, will use six codes and four episodes in the text that students may have underlined at least part of.

Article B. Because we started out with only six coding categories, we didn't need to collapse any of them to make the data more manageable.

Out of the 19 sentences, nine were underlined by at least 25% of the students in at least one grade level and 25% of the students who either do or don't believe that vaccines cause autism: 3, 4, 5, 6, 12, 13, 14, 17, and 18.

Unfortunately, it proved difficult to collapse these any further into episodes as we could for Article A. Although they fall into three distinct groups in terms of where in the article they are, those three groups weren't as cleanly distinct in terms of content as the episodes in Article A; most of the sentences talked the study that was the main topic of the article (3, 4, 5, 6, 12, 17, and 18), 13 and 14 discussed mechanism but gave two different opinions on it, and 12 and 13 were the only ones supporting the idea that vaccines do cause autism. This was reflected statistically: a factor analysis did not give clearly distinct factor loadings, nor were the sentences correlated with each other in any distinct groups (for example, 3 was significantly correlated with 4, 5, 12, and 13; but 5 was significantly correlated with 3, 4, 13, and 14).

Results

Note that, because our coding categories were not mutually exclusive, nor were which sentences students underlined, we were forced to limit ourselves to chi-squared tests. Because this results in more individual tests than other statistical methods would require, we used an alpha of .01 to determine significance.

What do students talk about when asked whether they believe vaccines cause autism and why?

Of the students who read Article A, 56.5% believed that vaccines caused autism while 43.5% did not, with no significant differences between grades.

Table 16: Percent of students who read Article A in each coding category

Anecdote	Correlation	Research	Coincidence	Authority	Mechanism
51%	24%	23%	16%	31%	49%

Table 16 shows what students talk about when explaining why they believe this. They are most likely to mention the anecdotal evidence in the article or a possible mechanism behind autism, whether the vaccine or otherwise, with about half of all students talking about one or both of these. 31% of students cited an authority figure such as a doctor or scientist, 24% talked about a correlation between vaccines and autism, and only 23% of students specifically mentioned any of the scientific research cited in the article. 16% of students cited the fact that any perceived correlation is or could be a coincidence.

The only significant difference between grades overall is those who mention Coincidence, as seen in Table 17.

Table 17: Percent of students who read Article A in each grade mentioning each code

	Anecdote	Correlation	Research	Coincidence	Authority	Mechanism
6th	49%	28%	15%	8%	24%	53%
7th	50%	13%	34%	18%	40%	42%
8th	56%	31%	31%	31%	36%	44%

Students are significantly more likely to mention this as they get older, which might indicate that older students have a better understanding of the difference between correlation and causation. Seventh and eighth graders were also twice as likely as sixth graders to mention the scientific research cited in the article, though this did not reach statistical significance.

Table 18: Percent of students who read Article A who do or don't believe there is a link that mentioned each code

	Anecdote	Correlation	Research	Coincidence	Authority	Mechanism	Other
Don't Cause	22.7%	1.5%	43.9%	31.8%	43.9%	24.2%	43.9%
Do Cause	73.3%	43.0%	8.1%	3.5%	20.9%	67.4%	29.1%

As seen in Table 18, students who do not believe vaccines or measles cause autism are most likely to talk about the research, coincidence, and authority figures in their answers, and more likely to talk about those than students who do believe it. Those who do believe it are most likely to talk about the anecdotes, mechanism, and correlation, and more likely to talk about those than students who don't believe it. Chi squares show each of these comparisons to be significant at the .01 level.

Only 32.4% of students who read Article B believed that vaccines cause autism, significantly fewer than those who read Article A ($X2=17.7$, $p<.001$).

Table 19: Percent of students who read Article B in each coding category

Research	Mechanism	Personal Experience	More Evidence/Info	Discrediting	Other
58.9%	35.6%	8.9%	13.7%	5.5%	15.1%

Unlike students who read Article A, 58.9% of students who read Article B mentioned the research cited in the article. However 37% of all students who mentioned the research (22% of students overall) were also found to have misinterpreted that research.

35.6% mentioned a possible mechanism. 8.9% talked about their personal experiences, 13.7% talked about one side having "more evidence" or "more information" without specifically mentioning anything in the article to make it clear what they thought of as "evidence/information", and 5.5% talked about the parts of the article where one expert tries to discredit the other. 15.1% talked about other topics (none of which were mentioned by enough students to merit individual analysis).

Table 20: Percent of students who read Article B who do or don't believe there is a link in each coding category

	Research	Misinterpreted	Mechanism	Personal Exp	MoreEvInfo	Discrediting	Other
No	56.6%	6.1%	27.3%	11.1%	16.2%	5.1%	15.2%
Yes	63.8%	57.4%	53.2%	4.3%	8.5%	6.4%	14.9%

Similar percentages of students who say vaccines do or don't cause autism cited the research in the article, though slightly more who said yes cited it.

However, far more students who say yes were likely to misinterpret that research. While 56.6% of those who said No cited the research, only 6.1% misinterpreted that research; on the other hand, while 63.8% of those who said Yes cited the research, 57.4% misinterpreted it. This means that out of the 56 students who said No and mentioned the research, only 11% misinterpreted it; meanwhile, fully 90% of the 30 students who said Yes and mentioned the research misinterpreted that research.

So overall, 90% of students who mentioned the research gave an answer that agreed with their interpretation of the research.

Those who believed that vaccines do cause autism were also more likely to cite a mechanism behind autism, but less likely to draw on personal experience in their answer. They were also less likely to state that there was simply more evidence or information in favor of their beliefs, and about as likely to mention the discrediting information or something else in their answer.

There were no significant differences in codes by grade, nor was there a significant difference in how many students in each grade misinterpreted the research.

Looking across the two articles at all of the students, even though there are some codes only applicable to one of the two articles we can still see a few trends in what students were likely to say when asked whether they believed vaccines cause autism.

As shown in Table 21, across both articles, older students were significantly more likely believe that vaccines do not cause autism ($X2=9.1$, $p=.01$), but there was no effect of grade on mentions of research ($X2=6.1$, $p=.05$) or mechanism ($X2=1.6$, $p=.45$). These were the only two common codes that were used on both articles.

Table 21: Students in each grade who did not believe vaccines cause autism and who mention research or mechanism

	Don't cause	Research	Mechanism
6th	46.70%	34.70%	45.30%
7th	65.20%	44.90%	36.20%
8th	62.70%	50.60%	42.20%

Students who read Article B, which focused on the details of scientific research, were significantly more likely than those who read Article A, which focused mainly on an anecdotal story with only brief mentions of research, to come to the conclusion that vaccines do not cause autism. 43.5% of those who read A believe vaccines do not cause autism, compared to 67.6% of those who read Article B. Given that over half of the students who read B and concluded that vaccines do cause autism had misinterpreted the research, it's very likely that if they had understood the details of the research the difference would be even greater.

Those who read B were also significantly more likely to talk about the research in the article in their defense of their positions (60.1% vs 23.4%), and less likely to talk about mechanism (35.8% vs 48.7%).

What did students identify in the article as scientific evidence?

Table 22: Percent of students who read Article A who underlined each episode

Anecdote	Science1	Science2	Mechanism
58.8%	49.1%	56.1%	47.4%

As shown in Table 22, overall, students who read Article A were slightly more likely to underline sentences in the opening anecdote (58.8%) and the second science episode (56.1%) than the first science episode (49.1%) or the episode that dealt with anecdotal evidence about a mechanism (47.4%).

Table 23: Percent of students who read Article A and do or don't believe there is a link who underlined each episode

	Anecdote	Science1	Science2	Mechanism
Don't	38%	51%	70%	30%
Do	73%	48%	46%	60%

Those who don't believe that vaccines cause autism tended to underline the second science episode, while those who do tended to underline the anecdotes. All of these differences were significant except for the Science1 episode - students who do believe they cause autism were just as likely to see this as scientific evidence as those who do not.

There were no significant differences between which episodes students in different grades underlined.

As explained in the data reduction section, nine of the nineteen sentences in Article B reached the threshold of 25% or more of students who either do or don't believe that vaccines cause autism and 25% of students in at least one grade. In addition, we were not able to collapse these nine sentences into episodes as we did with Article A. Table 24 shows the overall percentage of students who underlined each sentence.

Table 24: Percent of students who read Article B who underlined each sentence.

3	4	5	6	12	13	14	17	18
53%	67%	59%	40%	53%	51%	28%	33%	32%

Five of the nine, over a quarter of all the sentences in the article, were underlined by more than half of all students, showing much less variation in responses than with Article A. With seven of these nine sentences referring to scientific research, 94% of all students who read Article B identified at least one sentence referring to scientific research as counting as scientific evidence.

There were no significant differences between what those who believed vaccines do vs don't cause autism underlined, nor were there significant differences between the grades.

Students who read Article B were significantly more likely than those who read Article A to underline at least one sentence that referred to scientific research (94% to 73%, $X2=16.7$, $p<.001$). There was no significant difference, however, between the two articles when it came to underlining at least one sentence that referred to mechanism ($X2=2.2$, $p=.14$).

What relationships are there between what students wrote about and what they identified as scientific evidence?

Do students believe they are basing their decision on scientific evidence? If so, then we should see that, no matter what type of information they used to support their answers, they underlined sentences with the same type of information as being scientific evidence. That may not be the only thing they see as scientific evidence, but as long as they think it counts, then they thought they'd decided based on their understanding of scientific evidence.

The students who read Article A showed many significant relationships between what they discussed in their answers and what they underlined in the article.

Table 25: Percent of students who read Article A and mentioned a code who then went on to underline at least one sentence in a given episode.

	Anecdote	Science1	Science2	Mechanism
Anecdote	71%	48%	42%	54%
Correlation	86%	48%	38%	41%
Research	50%	82%	82%	43%
Coincidence	45%	85%	85%	40%
Authority	47%	50%	77%	47%
Mechanism	67%	49%	51%	66%

As you can see in Table 25, students who talked about the anecdotes, the correlation between autism and vaccines, and the mechanism by which autism is triggered are more likely to underline the anecdote than those who talked about the research, the timing of autism being a coincidence, or who cited authority figures.

Those who talked about research and coincidence were more likely than other students who underline the first science episode, and those students plus those who cited authority figures were more likely to underline the second.

Students who talked about mechanism were more likely than any other group to underline something in the episode discussing mechanism.

For each of these, we ran a chi squared test comparing those who did or did not get the code with those who did or did not underline a sentence in the episode. These are shown in Table 26.

Table 26: Significance testing of relationship between answers and underlines for students who read Article A

		Anecdote		Science 1		Science 2		Mechanism	
		No	Yes	No	Yes	No	Yes	No	Yes
Anecdote	No	55%	46%	49%	51%	29%	71%	60%	40%
	Yes	29%	72%	53%	48%	58%	42%	46%	54%
	X2	7.8 *		0.13		9.4*		2.3	
Correlation	No	51%	49%	51%	49%	38%	62%	51%	49%
	Yes	14%	86%	52%	48%	62%	38%	59%	41%
	X2	12.1*		0.01		5.2		0.56	
Research	No	38%	62%	62%	38%	52%	48%	51%	49%
	Yes	50%	50%	18%	82%	18%	82%	57%	43%
	X2	1.2		16.2*		10.2*		0.3	
Coincidence	No	38%	62%	59%	42%	50%	50%	51%	49%
	Yes	55%	45%	15%	85%	15%	85%	60%	40%
	X2	1.9		12.5*		8.2*		0.53	
Authority	No	36%	64%	51%	49%	53%	47%	53%	47%
	Yes	53%	47%	50%	50%	24%	76%	53%	47%
	X2	2.70		0.02		8.1*		0.00	
Mechanism	No	49%	51%	51%	49%	39%	61%	70%	30%
	Yes	33%	67%	51%	49%	49%	51%	35%	65%
	X2	3.20		0.00		1.20		13.9*	

Those who talked about the anecdote were more likely to underline it and less likely to underline Science2 than those who did not; those who talked about the correlation between autism and vaccines were more likely to underline the anecdote as well; those who talked about the research or that the timing was a coincidence were more likely to underline Science1 or Science2 than those who did not; those who cited an authority figure were more likely to underline Science2 than those who did not, and those who discussed the mechanism were more likely to underline the section describing mechanism.

Students who read Article B, on the other hand, did not tend to show significant relationships between their answers and what they underlined. As seen in Table 27, the only relationship that was significant at the .01 level was between students who talked about Mechanism and who underlined Sentence 14, which does talk about Mechanism.

Table 27: Significance testing of relationship between answers and underlines for students who read Article B

		Sentence 14	
		No	Yes
Mechanism	No	80.6%	19.4%
	Yes	57.1%	42.9%

While Article A and Article B are structured very differently and gave very different results in terms of the relationship between what students wrote and what they underlined, we can do a limited analysis across the two articles by looking at students who underlined at least one sentence referring to scientific research in either article and those who underlined at least one sentence referring to mechanism in either article.

Across the two forms, what students talked about in their answers had a significant effect on what they then identified as scientific evidence in the article they read.

Table 28: Significance testing of relationship between coding and underlines across both articles

		Underlined Mechanism		Underlined Research	
		No	Yes	No	Yes
Research	No	46%	54%	23%	77%
	Yes	50%	50%	9%	91%
	X2		0.29		7.2*
Mechanism	No	57%	43%	15%	86%
	Yes	34%	66%	20%	80%
	X2		11.1*		1.2

As Table 28 shows, those who talked about research were significantly more likely to then underline research in the article, but no more or less likely to underline mechanism, than those who did not talk about research. And those who talked about mechanism were more likely to then underline sentences referring to mechanism in the article, but no more or less likely to underline research.

This supports the idea that, overall, students believed themselves to be basing their answers on "scientific evidence."

Conclusions

We asked three research questions:

When reading a news article about an SSI, what do middle school students find most convincing?

How do they make use use of their understanding of scientific evidence when reasoning about the SSI in the article?

How do the journalistic features of the articles affect students' thinking?

The differences we found between the two articles were striking enough that it is difficult to discuss either of the first two questions without simultaneously addressing the third. Therefore, we will organize our conclusions around each of the first two questions as modified by the third.

When reading a news article about an SSI, what do students find most convincing and how do the journalistic features of the articles affect this?

Students given the article that was more balanced between the two viewpoints were also significantly less likely to come to the conclusion that vaccines cause autism than those who read the article that was biased in favor of a link. And of those who read the balanced article and thought that there was a link, over half of them misinterpreted the research presented, which would tip the balance of the article strongly in favor of the link. This indicates that it is not only the presence of both sides, as studied in Clarke (2008) and Dixon and Clarke (2013), that makes a difference, but how those sides are presented that makes a difference.

When given an article that uses multiple types of justifications to support the two positions—in this case, personal experiences, scientific research, statements from authority figures, and plausible mechanisms—students vary widely in which of these they use to support their opinion, all of them being mentioned by at least one in five students. Significantly more cite the anecdote and mechanism than the research or authority figures.

However, when given an article that presents fewer types of justifications—mostly scientific research, with only a few mentions of mechanism or statements by authority figures that aren't also about the research—the majority of students do talk about the research, with a third still mentioning mechanism. No other specific justification is given by more than 15% of students.

The patterns we've seen here indicate that, while most students do indicate that their opinion was influenced by something in the article, students only tend to find the research cited to be the most convincing justification when they are given few other options.

However, we cannot simply say that whichever justifications are given more attention are more likely influence students. In Article A, the anecdote is given 50% more words than the research, so it would make sense that it is mentioned far more often by students. However, this doesn't hold up when looking at mechanism. Taking every sentence that mentions any possible mechanism (including any that mention mercury in any capacity), only nine sentences, totaling 230 words, are devoted to mechanism. That's slightly less than the 14 sentences and 273 words talking about scientific research. (Note that there is overlap: most mentions of mechanism come during discussion of either the anecdote or the research.) Yet students were more than twice as likely to talk about mechanism when justifying their response to the issue.

The relationship between mechanism and evidence in students' minds is a complicated one. First, a plausible mechanism is an important part of a scientific explanation, and mechanistic reasoning is an important part of what distinguishes science from other ways of thinking (Russ, Scherr, Hammer, & Mikeska, 2008). We see justifying a claim through a plausible mechanism as preferable, for example, to justifying it purely based on a single anecdote.

But when the available empirical data conflicts with a suggested mechanism, as it does in Article A, in science it is the mechanism that must be changed to fit the data. Research has shown repeatedly that this can be difficult for children and adolescents. Young children appear to not even differentiate between evidence and mechanism, while older children and even adults see the difference but often find mechanistic explanations more persuasive than evidence (Kuhn, 2001). Sandoval and Cam (2010) found that while children preferred data as an explanation over appeals to authority, they gave equal weight to data and plausible mechanisms.

Given this general preference for mechanism over data, it is unsurprising that when the article gives the two nearly the same amount of space, middle school students are more likely to be persuaded by the mechanism over the data. Article B devotes only eight sentences, 191 words, to mechanism, versus eleven sentences and 272 words about scientific research (again, there is some overlap)—42% more words about research than mechanism, versus only 18% more in Article A.

Even with that increased difference, 36% of students who read Article B talked about mechanism in their answers.

It appears that while the amount of space given to a particular position or justification matters, it's not simply a matter of "whatever the article talks about more, people will talk about more." In order for students to be persuaded by the scientific evidence, it has to be given enough focus to overcome their general tendency to find mechanism more persuasive. Article A both devotes more space to the anecdote and couches it in a narrative, and it was still mentioned only about as much as mechanism, implying that anecdotal evidence does not fare much better than scientific evidence against mechanism.

It is important to note that in Article A, the only mechanisms mentioned in the article were mechanisms by which vaccines could cause autism, such as mercury or leaving traces of the virus in the body. Article B, on the other hand, included alternative mechanisms that could account for, if not autism itself, the rise in autism rates, such as increased awareness and diagnosis. This may partially account for the fact that more students who read Article B believed there was no link between vaccines and autism; however, given the fact that over half of the students who read Article B and concluded that there is a link misinterpreted the research, it is impossible to know where the credit for that truly lies.

How do they make use of their understanding of scientific evidence when reasoning about the SSI, and how do the journalistic features of the articles affect this?

According to our definition of "scientific evidence," only 23% of the students who read Article A and 59% of the students who read Article B based their conclusions on scientific evidence.

As we just discussed, they based their conclusions on mechanism at rates that outstripped the articles' focus on mechanism.

They had similar tendencies when underlining what they classified as scientific evidence in the articles. Those who read Article A were about equally likely to to underline one or more sentences in the anecdote, the two sections focusing on scientific research and data, or the section focusing on mechanism. Most of the most commonly-underlined sentences in Article B were focused on scientific research, but several of those do mention mercury. Out of the nine sentences, three talk about the research with no mention of mercury, four are focused on the research but do mention mercury, and two are entirely focused on mechanism with no mention of research.

These suggest that these students do not necessarily define "scientific evidence" the way that we would, or the way that we would hope that our educational system prepares adults to.

This is supported by the relationship between what they used to justify their answer and what they later underlined as scientific evidence.

Across the two articles, those who used research in their justification were more likely than those who didn't to classify at least one sentence focused on research as scientific evidence, and those who used mechanism in their justification were more likely than those who didn't to classify at least one sentence focused on mechanism as scientific evidence. This indicates that these students believed themselves to be basing their arguments on scientific evidence, even when we would not have coded it that way.

However, this is strongly influenced by which article students read. There were many significant correlations between what students talked about and what they later underlined in Article A. Students who talked about anecdote were more likely to underline the anecdote and less likely to under line scientific research, students who talked about research were more likely to underline research, and students who talked about mechanism were more likely to underline mechanism. In addition, those who mentioned the fact that children tend to develop autism right after getting the vaccine, used as evidence as part of the anecdote, were more likely to underline the anecdote, while those who explicitly dismissed this as a coincidence, a position put forth by a scientist and supported by scientific data in the article, were more likely to underline not just that bit of scientific data but also the other section talking about scientific research.

Overall, when reading an article that presents many types of justifications for the two sides, students were likely to believe that whichever of those they used in their own argument counted as scientific evidence.

The students who read Article B, which varied less in its justifications, did not show the same relationships. We found no significant relationship between what students who read Article B talked about in their justifications and what sentences they classified as scientific research, with only one exception: of those who talked about mechanism, 43% underlined Sentence 14, which was focused on mechanism, while only 19% of those who did not talk about mechanism underlined this sentence. This one significant relationship, while suggestive, is not enough to say that students who read Article B tended to classify what they wrote about as scientific evidence overall.

Article B devotes 57% of its words to discussion of scientific research, whereas Article A only gives research 35% of its words. There is no anecdote at all, and only 20% of the article is about mechanism and not also scientific research. The remaining 23% are devoted to discrediting the scientist who ran the study, which means that they are, indirectly, about the research as well. As we've already mentioned, only two of the nine most-underlined sentences are not about scientific research, and both of those are about mechanism; none of the discrediting sentences were underlined by even 10% of students.

Not only that, but most of the sentences discussing scientific research in Article B are about a single study, and the discrediting sentences are also in reference to that one study. Even three of the sentences that are about mechanism are proposing mechanisms to either support or refute the study's findings. Article A discusses several studies and sources of data, as well as at least two separate anecdotal cases. Thus both the content of the article and the most-underlined sentences were much more homogenous in Article B than Article A.

In Article A, all of the justifications presented against a link were scientific research or data, or a statement from an authority figure that references scientific research, while none of the justifications for the link were. The justifications for a link varied widely: personal experiences, plausible mechanisms, anecdotal evidence that sounds scientific, and statements from authority figures. Yet Article A gives nearly twice as much space to the pro-link side of the argument, a bias which could influence readers in and of itself. Thus it creates a false equivalency between the scientific research and all of these other justifications; if they are equally valid as justifications in an argument about a scientific topic, that could lead a reader to conclude that they are equally valid as scientific evidence.

In Article B, most of the justifications on both sides of the argument were focused on a particular scientific study, whether discussing the actual results, possible mechanism to account for those results, or the qualifications of the man who ran it. Making this study so central to the debate makes it clearer that scientific research is the most valid type of evidence; even those who found the mechanisms most persuasive may be able to see that those mechanisms are discussed in service of the research rather than instead of it, and thus be less likely to believe them to count as scientific evidence.

This effect supports an epistemological resources approach to beliefs about science, as it shows the context dependence not only of students' reasoning, but their beliefs about their own reasoning. Article A's presentation of several other justifications as equivalent to scientific research led to the activation of the scientific evidence resource inappropriately more often than did the more focused article.

Discussion

This study highlights the complexity involved in learning about SSIs from news media sources, as most people do. The two articles chosen represented two very different ways of approaching the same issue, from the extent to which they "balanced" the two sides of a debate which has scientific support for only one of the sides, to the way they framed the issue, to the style of prose used.

These differences had such an impact that there was little that we could say held true across both articles. Which article students read had a bigger impact on their thinking, it seems, than what grade they are in, which only led to significant differences in which side of the debate they believed.

Our results show that news articles must focus heavily on research to overcome students' default bias toward the persuasiveness of mechanism. They support Dixon and Clarke's (2013) findings that false balance can lead readers to incorrect conclusions, but show that "balance" is more complex than simply whether or not each side is mentioned. They show that a reliance on justifications other than scientific research can influence not only which justifications readers find most convincing, but which justifications they believe count as scientific evidence.

This, in particular, supports the idea that our beliefs about science itself are not stable, but highly context-dependent. They can be influenced by a single news article. The epistemological resources framework posits that the more often an ER is activated in a particular context, the more likely it is to be activated by that context in the future, and that the more often multiple ERs are activated as a group, or frame, the more likely that a similar context will activate not just one ER but all ERs in that frame. It is possible that reading many news articles that lead to the inappropriate association of the scientific evidence ER with mechanism or anecdotes could make it more likely that, in the future, an article that discusses these but does not equate them with scientific research will still activate this ER. In other words, reading too many articles like Article A could theoretically lead to a reader having more trouble correctly differentiating the scientific evidence from other justifications even when reading an article more like Article B. Obviously, far more research is needed before such a conclusion can be reached, but this study suggests that the epistemological resources framework is a productive way to look at this issue and that this would be a fruitful direction for future research.

This study has many limitations. The desire to use authentic news articles meant that we could not control the content of those articles tightly. While in some cases we were able to make logical guesses about which differences in the articles might account for differences in the results, we cannot make definitive causal links between any one feature with any one result. We did not employ any sort of pre-assessment, so we do not know whether the students were familiar with this debate before reading these articles or what their opinions on it were. As the two articles were randomly distributed within classes, we have no reason to believe that there should be any systematic differences between the two groups that would fully account for the differences we see between the articles, but there could be an interaction between prior knowledge and the articles. Finally, as our sample was taken from two schools that are both implementing the iQWST curriculum that emphasizes evidence and mechanistic reasoning and the relationship between the two, it is possible that students who have not been taught from this type of curriculum might show even less stability in their conceptions of scientific evidence.

Many researchers now promote the use of SSIs as a context within which to teach science in the classroom (Tal & Kedmi, 2006; Sadler, Barab, & Scott, 2007; Zeidler, Sadler, Applebaum, & Callahan, 2009), as they are both more similar to the contexts in which students will be encountering science as adults and complex environments that promote multiple types of critical thinking. This study implies that while real-world SSIs may be useful educational contexts, we should be careful when extending that to the use of real news articles. The use of real news articles would increase the similarity to real-world contexts, but journalistic norms may make learning from these more complicated than from texts written for science education. Educators must take into account this added complexity and be prepared to help students navigate textual features that could lead to confusion. If attention is paid to this, the inclusion of real news articles could help make students even more prepared for decisions they may have to make in the future that require the evaluation of scientific evidence and arguments.

Chapter 4:

Using the Results of the Epistemology Assessment to Predict Behavior on the News Assessment

In both of the first two studies, I emphasized the fact that science education purports to prepare students to engage with science in the real world, and that we need to find ways to assess whether this is what we are, in fact, doing.

In Study 1, I presented an assessment made up of written performance tasks. These were designed to allow us to see what epistemological resources students activated to complete the tasks, and I found that the tasks were at least successful in provoking behavior that showed a difference in ERs activated between sixth graders and eighth graders.

In Study 2, I gave students actual news articles about a socioscientific issue and asked them to answer the central question addressed by these articles (Do vaccines cause autism?), and then asked them to underline the parts of the article that they considered to be scientific evidence. I found that the journalistic choices made in the two articles had a major impact not only on how they reasoned about the issue, but also what they identified as scientific evidence and the relationship between the two.

In Study 3, I will go back to the original question of how to assess whether science education truly prepares students to engage with science in the real world, by investigating whether the epistemology assessment from Study 1 tells us anything about how students engage with the news articles in Study 2.

Theoretical Framework

The theoretical framework outlined in Study 1 (and illustrated in Figures 1 and 2), rooted in the epistemological resources view that epistemological knowledge is highly context-dependent and implicit, suggests that if our goal is to predict students' behavior outside of school, our assessment context must simulate that context in ways that will prompt students to activate the same epistemological frames and resources that they will in real life. One way in which we tried to do this was by grounding the assessment in the evaluation of scientific knowledge products (arguments, explanations, and models), because most students will need to evaluate these more than they will need to create them. But Study 1 did not address the question of whether these were successful in activating the same frames that students would activate in more realistic situations.

However, people rarely encounter these knowledge products in isolation, or in a form straight from the scientists who constructed them. Instead, they usually encounter them as a part of a larger socioscientific issue or debate, and filtered through the lens of the mass media. Study 2 showed us that journalistic choices are an important part of this context, and can have a significant impact on people's reasoning. Students who read Article B were more likely to come to a correct conclusion about the scientific question and to attend to scientific research. Students who read Article A, on the other hand, were more likely to use anecdotal evidence or plausible mechanisms in their reasoning, and to believe that whatever they had used in their reasoning counted as scientific evidence, possibly due to journalistic choices that implied a false equivalency between the actual research and the anecdote and mechanism.

According to my theoretical framework, if I was successful in simulating the features of real-world contexts that determine students' epistemological framing, then the results of my epistemological assessment should predict students' behavior when presented with a news article about science. But given the impact of the specific contextual features of each article on how students framed the issue, the epistemology assessment may tell us different things about how students will reason about each article.

Specifically, science educators should especially want to equip students to avoid the traps set by articles like A to distract them from the actual research. It would be particularly useful if our epistemology assessment could tell us anything about students' epistemological framing that would help us predict students' susceptibility to articles like this.

Research Questions:

1. Does the epistemological assessment predict anything about the students' reasoning about the news articles?

2. Does anything we can learn from the epistemological assessment predict whether students will be more or less susceptible to the false equivalency presented in Article A?

Methods

Participants

Participants are the same pool of participants as Study 1 and Study 2, from two midwestern suburban middle schools. We included all students who completed both the epistemology and news assessments, resulting in a sample of 232 students made up of 141 sixth graders and 91 eighth graders. This included 109 males and 123 females, with similar ratios in each of the two grades.

Because we used data from fall of the sixth grade and spring of the eighth grade for Study 1, and the news assessment for Study 2 was given mid-year, sixth graders were taking the news assessment a couple of months after the epistemology assessment for Study 1 while eighth graders were taking it a couple of months before the assessment in Study 1. Eighth graders had taken the same assessment in the fall, though that data was not included in the analysis for Study 1, so all students had been previously exposed to the epistemology assessment when they took the news assessment.

Coding

For the most part, we used the coding that had already been completed for Study 1 and Study 2. However, the way we coded answers in Study 1 was purposefully broad, looking to see if a student made use of a particular ER at all. This is not necessarily the best way to use this assessment to predict the results of Study 2. In particular, when coding the first question, where students were asked to evaluate an argument between two scientists, we coded for evidence simply by looking at whether and how many times the student evaluated an argument according to the quantity or quality of the evidence used in that argument—regardless of what that evidence was or what the student's evaluation of it was. This means that a student who talked about Anne's use of a single anecdotal case favorably were counted in the same code as students who talked about the use of this negatively or who talked about the use of the data from their study favorably.

In Study 2, we were interested in *what*, specifically, the students counted as scientific evidence and what types of information they found valuable and useful as a basis for their decision-making. From a theoretical standpoint, we are less interested in *whether* they activated the "Evidence" ER and more interested in *what type of information* they attached that ER to. There is no reason to think that the way we coded for evidence in Study 1 is the best way to use the results of that assessment to predict performance on Study 2.

Therefore, we went back and added four codes with respect to this question and students' handling of evidence. In each case, we counted the number of times students expressed each idea, explicitly or implicitly, when evaluating the strength and weaknesses of the two arguments or explaining why one argument was better:

1. Anne's anecdote is useful and valuable evidence. (Anecdote Good Score)

2. Anne's anecdote is not useful or valuable, or less so than other types of evidence. (Anecdote Bad Score)

3. The data from the experiment is useful and valuable evidence. (Cherry-Picking Score)

4. Peter's cherry-picking of the data is not a good way to use that evidence. (Data Good Score)

Students may express one of these ideas explicitly by saying something like "Peter [had the better argument], because he explained why people shouldn't take the pill using the chart." This would be counted toward category 3, as "the chart" refers to the data from the experiment. They may also express one of these ideas implicitly, by, for example, listing something as a strength or weakness of one of the articles. For example, one student listed "She talks with someone who took the pill" as a strength of Anne's argument and "She doesn't have enough support from the data" as a weakness of it. These statements would be counted toward category 1 and 3 respectively, as listing the first one as a strength implies that the anecdote is valuable evidence, while listing *not* using the study data as a weakness implies that the data is valuable evidence.

As there were five places for students to make distinct statements (strengths and weaknesses of Anne's argument, strengths and weaknesses of Peter's argument, who had the better argument), they could theoretically get up to a 5 in each of these categories, though in practice these counts rarely went about 3.

Analysis

Our outcome variables were questions with yes or no answers, so we ran binary logistic regressions to look for relationships between their answers in Study 1 and these questions.

For each outcome variable, we tested a model containing as predictor variables all of the scores from the epistemology assessment (Evidence, Cause and Effect, Amount of Information, Correctness, and Causal Coherence) and three of the four new coding categories above, along with Grade and Article Read. We found that the Cherry-Picking Score was highly correlated with the Data Good Score; the Data Good Score had a higher range and more variance, and so should be the more sensitive measure, so we dropped the Cherry-Picking Score. The Causal Coherence score was multiplied by 10 to give a score from 0-10 rather than 0-1, to make interpretation of regression results easier. If the model's Chi Squared statistic wasn't significant, but one or more variables was significant, we used a backwards stepwise regression on the Wald statistic to remove variables until we found a significant model, then reran that model separately to ensure that the stepwise results were valid. We are reporting the results of that final model, with all variance reported as Nagelkerke R^2.

Hypotheses

Along with whether students believed that vaccines cause autism after reading the article, we looked at the three most common types of information students in Study 2 used to explain their answers and underlined as scientific evidence: scientific research, anecdotal evidence, and plausible mechanism. For each of these we asked whether the student mentioned it in their answer and whether they underlined it (though anecdotal evidence was only relevant for those who read Article A). In Study 2, we also found that students who read Article A were likely to underline the same type of information that they used to support their answer, while students who read Article B were not. To explore predictors of this interaction, we created variables that contained a 1/Yes if the student *both* mentioned *and* underlined a particular type of information, and a 0/No if they only did one or neither.

Based on previous results, we hypothesize that:

1. Grade will be a predictor of their answer to the vaccine/autism question and nothing else, as there were few significant differences between the grades in Study 2.

2. Evidence Score in and of itself will not predict any results for reasons stated in the Coding section, but Anecdote Good Score, Anecdote Bad Score, Cherry Picking Bad Score, and Data Good Score will. Anecdote Good Score will be associated with more reliance on the anecdote and less on research, while Anecdote Bad Score, Cherry Picking Bad Score, and Data Good Score will be associated with less reliance on the anecdote and more on research. There is no reason to expect these to be associated with mechanism, given that the question they are from did not address mechanism.

3. Correctness Score and Amount of Information Score will be associated with less reliance on research and mechanism. This is because these codes were only used if the student talked about these concepts *without* relating them to evidence or mechanism, so it is possible that those students are unlikely to pay attention to these specifically over other types of information.

4. Cause and Effect Score and Causal Coherence may be positively associated with mentioning mechanism in their answer, but not with identifying it as scientific evidence. This is because these scores reflect a higher reliance on Cause and Effect *in situations where it is relevant*. While plausible mechanism is one valid way to support a claim, it should not be considered scientific evidence, so if the students understand when it is and isn't relevant they shouldn't underline it but may talk about it.

5. Article Read will predict whether they believe vaccines cause autism, mentioning and underlining research, and mentioning (but not underlining) mechanism, as well as doing both of these at once.

Results

Across both articles

These results are summarized in Table 29.

Did the student believe that vaccines cause autism?

The significant model (ChiSq=27.25, $p<.01$) accounted for 19% of the variance and predicted 67% of the answers to this question, with two significant predictors, Article Read and Data Good Score. Students who read Article B were only 27.4% as likely to believe that vaccines cause autism as those who read Article A (Wald=14.0, $p<.001$). For every additional point on Data Good Score, students were 64.9% as likely to believe this (Wald=4.6, $p<.05$).

Did the student use the research to explain their answer?

The significant model (ChiSq=60.10, $p<.001$) accounted for 39% of the variance and predicted 74% of the answers to this question, with three significant predictors, Article Read, Cause and Effect Score, and Data Good Score. Students who read Article B were 9.6 times as likely to mention research as those who read Article A (Wald=30.1, $p<.001$). For every additional point on Cause and Effect Score, students were 1.2x as likely to mention research (Wald=4.7, $p<.05$), and for every additional point on Data Good Score, students were 1.7 times as likely (Wald=5.9, $p<.05$).

Did the student use mechanism to explain their answer?

There were no models that significantly predicted this answer.

Did the student use the anecdote to explain their answer? (Applies to those who read Article A only.)

There were no models that significantly predicted this answer.

Did the student identify at least one sentence about scientific research as scientific evidence?

The significant model (ChiSq=25.45, $p<.01$) accounted for 32% of the variance and predicted 87% of the answers to this question, with four significant predictors, Article Read, Amount of Information Score, Causal Coherence, and Data Good Score. Students who read Article B were 7.0 times as likely to underline research as those who read Article A (Wald=7.4, $p<.01$). For every additional point on Amount of Information Score, students were 1.6x as likely to mention research (Wald=4.5, $p<.05$), for every additional .1 point on Causal Coherence, students were 1.2 times as likely (Wald=4.2, $p<.05$), and for every additional point on Data Good Score, students were 2.9 times as likely (Wald=5.2, $p<.05$).

Did the student identify at least one sentence about mechanism as scientific evidence?

There were no models that significantly predicted this answer.

Did the student identify at least one sentence about the anecdote as scientific evidence? (Applies to those who read Article A only.)

There were no models that significantly predicted this answer.

Did the student both mention scientific research <u>and</u> identify at least one sentence about the research as scientific evidence?

The significant model (ChiSq=48.40, $p<.001$) accounted for 42% of the variance and predicted 75% of the answers to this question, with two significant predictors, Article Read and Data Good Score. Students who read Article B were 10.9 times as likely to mention research *and* underline it as those who read Article A (Wald=23.8, p<.001). For every additional point on Data Good Score, students were 2.4 times as likely (Wald=11.0, p<.001).

Did the student both mention mechanism and identify at least one sentence about mechanism as scientific evidence?

The significant model (ChiSq=9.83, p<.05) accounted for 10.4% of the variance and predicted 79% of the answers to this question, with one significant predictor, Cause and Effect Score. For each additional point on Cause and Effect Score, students were 1.3 times as likely to mention mechanism *and* underline it (Wald=5.5, p<.05).

Did the student both mention anecdote and identify at least one sentence about the anecdote as scientific evidence? (Applies to those who read Article A only.)

The significant model (ChiSq=11.81, p<.05) accounted for 21% of the variance and predicted 66% of the answers to this question, with one significant predictor, Data Good Score. For every additional point on Data Good Score, students were 48.8% as likely to both mention anecdote *and* underline it (Wald=5.0, p<.05).

Table 29: Significant results of binomial logistic regressions.

				Significant Predictors			
	ChiSq	R^2	Predicted	1	2	3	4
Answer	27.3**	.19	67%	Article	Data Gd.	-	-
B				-1.29	-0.432		
Exp(B)				0.274	0.649		
Wald (p)				14.0**	4.6*		
Mentioned							
Research	60.1***	0.39	74%	Article	Cause/Eff.	Data Gd.	-
B				2.27	0.207	0.544	

Exp(B)				9.6	1.23	1.72	
Wald (p)				30.1***	4.7*	5.9*	
Underline							
Research	25.4**	.32	87%	Article	Amt Info	Causal Co.	Data Gd
. B				1.95	0.450	0.187	1.07
Exp(B)				7.04	1.57	1.21	2.91
Wald (p)				7.4**	4.5*	4.2*	5.7*
Both							
Research	48.4***	0.42	75%	Article	Data Gd.	-	-
B				2.39	0.861		
Exp(B)				10.9	2.37		
Wald (p)				23.8***	11.0***		
Mechanism	9.8*	.10	79%	Cause/Eff.	-	-	-
B				0.238			
Exp(B)				1.27			
Wald (p)				5.5*			
Anecdote	11.8*	0.21	66%	Data Gd.	-	-	-
B				-0.716			
Exp(B)				0.488			
Wald (p)				5.0*			

**p<.05 **p<.01 ***p<.001*

Between Articles

What were the differences between students who read the two articles? We reran these models on the two articles separately to find out whether the article read affected how the epistemology assessment results predicted the news assessment results.(Article Read will always be dropped as a predictor, since each analysis was run on only one article. Comparisons cannot be made for variables involving the anecdote.) These results are summarized in Table 30.

Did the student believe that vaccines cause autism?

For both articles individually, there were no significant models.

Did the student use the research to explain their answer?

For students who read Article A, no models significantly predicted this answer.

For students who read Article B, the significant model (ChiSq=16.11 p<.05) accounted for 24% of the variance and predicted 67% of the answers with one significant predictor, Data Good Score. For every additional point on Data Good Score, students were 1.9x as likely to talk about research in their answer (Wald=4.2, p<.05).

Did the student use mechanism to explain their answer?

For students who read Article A, the significant model (ChiSq=12.89 p<.05) accounted for 17% of the variance and predicted 62% of the answers with one significant predictor, Anecdote Good Score. For every additional point on Anecdote Good Score, students were 39.4% as likely to talk about mechanism in their answer (Wald=6.3, p<.05).

For students who read Article B, no models significantly predicted this answer.

Did the student identify at least one sentence about scientific research as scientific evidence?

For students who read Article A, the significant model (ChiSq=16.75 p<.05) accounted for 36% of the variance and predicted 79% of the answers with one significant predictor, Data Good Score. For every additional point on Data Good Score, students were 5.6 times as likely to underline research (Wald=5.9, p<.05).

For students who read Article B, the significant model (ChiSq=15.00 p<.01) accounted for 49% of the variance and predicted 96% of the answers with one significant predictor, Anecdote Bad Score. For every additional point on Anecdote Bad Score, students were 9.9 times as likely to talk about mechanism in their answer (Wald=4.0, p<.05).

Did the student identify at least one sentence about mechanism as scientific evidence?

For students who read Article A, the significant model (ChiSq=19.93 $p<.05$) accounted for 35% of the variance and predicted 73% of the answers with two significant predictors, Anecdote Bad Score and Data Good Score. For every additional point on Anecdote Bad Score, students were 5.8 times as likely to underline mechanism (Wald=6.7, $p<.01$). For every additional point on Data Good Score, students were 2.2 times as likely to underline mechanism (Wald=4.2, $p<.05$).

For students who read Article B, no models significantly predicted this answer.

Did the student both mention scientific research <u>and</u> identify at least one sentence about the research as scientific evidence?

For students who read Article A, the significant model (ChiSq=14.69 $p<.05$) accounted for 33% of the variance and predicted 85% of the answers with one significant predictors Data Good Score. For every additional point on Data Good Score, students were 3.0 times as likely to mention research *and* underline it (Wald=7.0, $p<.01$).

For students who read Article B, no models significantly predicted this answer.

Did the student both mention mechanism <u>and</u> identify at least one sentence about mechanism as scientific evidence?

For students who read Article A, the significant model (ChiSq=19.93 $p<.05$) accounted for 35% of the variance and predicted 73% of the answers with four significant predictors, Grade, Amount of Information Score, Cause and Effect Score, and Anecdote Good Score. Eighth graders were 11% as likely as sixth graders to both mention mechanism *and* underline it (W=4.9 $p<.05$) For every additional point on Amount of Information Score, students were 1.9 times as likely to do both of these (Wald=5.4, $p<.05$), for every additional point on Cause and Effect Score students were 1.6

times as likely (Wald=4.2, p<.05), and for every additional point on Anecdote Good Score students were 19% as likely (Wald=5.8, p<.05).

For students who read Article B, the significant model (ChiSq=16.00 p<.05) accounted for 34% of the variance and predicted 79% of the answers with two significant predictors, Grade and Data Good Score. Eighth graders were 6.5 times as likely as sixth graders to both mention mechanism *and* underline it (Wald=4.1, p<.05). For every additional point on Data Good Score, students were 40% as likely to do both (Wald=4.0, p<.05).

Table 30: Significant results of binomial logistic regressions per article.

				Significant Predictors			
	ChiSq	R^2	Predicted	1	2	3	4
Mentioned							
Research							
Article A	-						
Article B	16.1*	0.24	67%	Data Gd.			
B				0.646			
Exp(B)				1.91			
Wald				4.2*			
Mechanism							
Article A	12.9*	0.17	62%	Anec Gd.			
B				-0.93			
Exp(B)				0.394			
Wald				6.3*			
Article B	-						
Underline							
Research							
Article A	16.7*	0.36	79%	Data Gd.			
B				1.72			
Exp(B)				5.57			
Wald				5.9*			
Article B	15.0**	0.49	96%	Anec Bad			
B				2.3			
Exp(B)				9.95			
Wald				4.0*			
Mechanism							
Article A	19.9*	0.35	72%	Anec Bad	Data Gd.		

B				1.75	0.81		
Exp(B)				5.78	2.25		
Wald				6.7**	4.2*		
Article B	-						
Both							
Research							
Article A	14.7*	0.33	85%	Data Gd.			
B				1.08			
Exp(B)				2.95			
Wald				7.0**			
Article B	-						
Mechanism							
Article A	24.2**	0.45	86%	Grade	Amt Info	Cause/Eff	Anec Gd.
B				-2.2	0.654	0.469	-1.67
Exp(B)				0.111	1.92	1.6	0.188
Wald				4.9*	5.4*	5.3*	5.8*
Article B	16.0*	0.34	79%	Grade	Data Gd.		
B				1.87	-0.918		
Exp(B)				6.48	0.399		
Wald				4.1*	4.0*		

**p<.05 **p<.01 ***p<.001*

Conclusions

As we saw in Study 2, which article students read had a major impact on their responses. Those who read Article B were seven to eleven times more likely to interact with the scientific research in the article than those who read Article A. Beyond that, it also changed which epistemology scores predicted their results in many categories.

Across both articles, Data Good predicted more results than any other epistemology score. The more times a student talked about the data as a valuable form of evidence in Question 1, the more likely they were to believe that vaccines *don't* cause autism, to mention or underline research or do both. Students were also more likely to base their answer on the research if they had higher Cause and Effect scores, and more likely to identify research in the article as scientific evidence if they had higher Amount of Information or Causal Coherence scores. They were more likely to both mention mechanism in their answer *and* identify it as scientific evidence if they had higher Cause and Effect scores.

Those who read Article B, which was focused on a single scientific study, were more likely to make use of the research overall. But within that group, those who had higher Data Good scores were even more likely to mention the research They were also more likely to both mention and underline mechanism, as were eighth graders over sixth graders. Those who talked about the fact that the anecdote in Question 1 was a poor type of evidence were also more likely to underline research as scientific evidence, if they read Article B.

Article A contained anecdotal evidence that B was lacking, and while none of the epistemology scores made a student more or less likely to mention or underline the anecdote, higher Data Good scores made them less likely to do both. Data Good scores also made them more likely to underline the research, or to mention *and* underline it.

Mechanism had many more predictors for those who read Article A than it did for B. Students who implied that the anecdote was valuable evidence in Question 1 (Anec Good Score) were less likely to to talk about mechanism and less likely to both talk about and underline mechanism. Those who implied that it was *not* valuable evidence (Anec Bad Score) were more likely to underline mechanism, as were those who had higher Data Good scores. Sixth graders were less likely to both talk about an underline mechanism, while those with higher Amount of Information and Cause and Effect scores were more likely to do both.

In terms of our individual hypotheses, some were supported and some were not.

1. Grade will be a predictor of their answer to the vaccine/autism question and nothing else. In the end, when the results of the epistemology assessment were taken into account, grade did not predict their answer at all. The only thing that it did predict was whether they both mentioned and underlined mechanism, and then it was in a different direction for each article.

2. Evidence Score will not predict any results. Anecdote Good Score will be associated with more reliance on the anecdote and less on research, while Anecdote Bad Score and Data Good Score will be associated with less reliance on the anecdote and more on research. As predicted, the Evidence Score was not a significant predictor in any model. Anecdote Good Score was not at all associated with either anecdote or research, though it did predict some use of mechanism in Article A. Anecdote Bad Score was slightly associated with use of research (underlining in Article B), but not at all with anecdote, and it also predicted attention to mechanism for Article A. Data Good Score was consistently associated with more reliance on the research, and was somewhat associated with less reliance on the anecdote.

3. Correctness Score and Amount of Information Score will be associated with less reliance on research and mechanism. Correctness was not a significant predictor of anything. Amount of Information was positively, rather than negatively, associated with both research and mechanism.

4. Cause and Effect Score and Causal Coherence may be positively associated with mentioning mechanism in their answer, but not with identifying it as scientific evidence. Cause and Effect was actually associated with the combination of both mentioning *and* underlining mechanism, along with mentioning research. Causal Coherence was only associated with underlining research, and nothing to do with mechanism.

5. Article Read will predict whether they believe vaccines cause autism, mentioning and underlining research, and mentioning (but not underlining) mechanism, as well as doing both of these at *once.* Article Read predicted all of these except the ones involving mechanism.

As expected, our answers to our two research questions are complex.

1. Does the epistemological assessment predict anything about the students' reasoning about the news articles?

Aside from which article students read, the Data Good scores were the most useful in predicting their behavior on the news assessment. While the Evidence score, which only recorded how much they relied on the idea of evidence at all in their answers, did not predict anything, when we break that down into what types of evidence they valued it is clear that those who indicate that the data from the experiment is valuable evidence are more likely to pay attention to the research in the articles. They are also less likely to both base their answer on anecdotal evidence and view that as scientific evidence.

Two measures of students' attention to cause and effect—the causal coherence of their explanations and the number of times they brought up cause and effect when evaluating explanations and models— predicted that students were more likely to either mention *or* underline the research or to both mention *and* underline mechanism. One possible interpretation of this is that, in general, an understanding of when it is important to pay attention to cause and effect also results in an understanding of when it is less important to pay attention to that than to things like research results and data. However, those students who focused on cause and effect on that assessment who *do* wind up paying attention to mechanism when making decisions might also be more likely to identify it as scientific evidence—implying that while some students with high Cause and Effect scores understand when it is and isn't appropriate to use, some are prone to just looking for it everywhere.

The only other score that predicted anything across both articles was Amount of Information; higher scores there indicated a higher likelihood of underlining the research in the articles as scientific evidence. The relationship there is unclear. When coding the epistemology assessment, we were careful to only use this code if the student was relying on the amount of information given in their evaluation *but not* the type of information—so this was given to students who said that a good answer had more "details" or "facts" but not one who said that the answer explained "how" or "why" something happened, or who said that the answer had more "evidence" or "data," for example. So it doesn't necessarily mean that they *aren't* paying attention to the salient information, just that they either don't know which type of information is more important *or* that they don't yet have the *language to explain* which information is important. Being asked to underline scientific evidence may help those students who know what they are looking for but might not be able to come up with evidence-related words themselves.

2. Does anything we can learn from the epistemological assessment predict whether students will be more or less susceptible to the false equivalency presented in Article A?

In general, the epistemology assessment predicted far less about the performance of students who read Article B than those who read Article A.

The content and structure of Article B, which focused on a single study, gave slightly more space to the side of the debate supported by the evidence, and did not rely on anecdotal evidence at all, seems to have influenced students' thinking more regardless of their performance on the epistemology assessment. The only information we can glean from the epistemology scores is that those who valued data were more likely to talk about research in their explanations and less likely to both mention mechanism *and* underline it, while those who said negative things about anecdotal evidence were more likely to underline the research in the article.

Students' responses to Article A, on the other hand, were predicted by a number of results on the epistemology assessment. Our concern here is that this article sets up a false equivalency, citing scientific research and data to support one side of the debate (no link between vaccines and autism) while using anecdotes and plausible mechanisms to support the other side, but giving the pro-link side twice as much space and using narrative structures to persuade the reader. This could give the impression that the pro-link side of the debate is the one more worth listening to, and thus the anecdotes and plausible mechanisms are equally valid evidence as the fourteen large studies that are glossed over in a few sentences.

While we would hope, of course, that students would attend to the scientific research even though it is given far less space, we also acknowledge that plausible mechanisms are an important part of scientific explanations and models. So while scientific evidence should take precedence over a conflicting mechanism, the mechanism is preferred as support for an argument to anecdotal evidence. This means that while, in general, we would prefer students talk about the research and identify it as evidence, which they generally do for Article B, if the structure of Article A were to influence their thinking we would at least prefer they attend to the mechanism over the anecdotal evidence.

Our data suggest that the epistemology assessment can give us some information about which students are more likely to be fooled by Article A. The only thing that predicts their attention to the research is the Data Good score, which was associated with underlining the research more as well as both talking about and underlining the research more. This was also the only thing that predicted anything about their attention to the anecdote, in that higher Data Good scores mean that students were less likely to both talk about and underline the anecdote. If our main concern is ensuring that students will pay attention to the scientific research over anecdotal evidence, our best predictor is how much they valued the scientific data when evaluating arguments.

However, the other parts of the assessment do give us insight into whether students are likely to focus on mechanism when reading this article. Though the mechanisms discussed in the article have been disproven by the scientific evidence given, reasoning based on those mechanisms would at least be more scientifically sound than reasoning based on the single anecdotal case discussed.

Students who read Article A were less likely to use mechanism in their reasoning the more they said positive things about the anecdotal evidence in Question 1, and more likely to underline the mechanism as scientific evidence the more they said negative things about the anecdotal evidence and positive things about the research data in that question. Both mentioning and underlining the mechanism together was associated with talking more about cause and effect and amount of information on the epistemology assessment, and students were less likely to do this the more they said positive things about the anecdote.

So while only the Data Good score gives us information about which students will prioritize the scientific research or give the anecdote less priority, several measures predict their focus on the mechanism in ways that they *do not* predict this behavior for students who read Article B. As with the overall results, it is unsurprising that students who focused on cause and effect in the epistemology assessment would continue to focus on it here, even if this indicates that they do not yet know which situations are appropriate to apply this idea to. And while we would hope that students' treatment of the anecdotal evidence and data in the epistemology assessment would predict their being more or likely to choose scientific research over the anecdote here, it is promising that at least it predicts their likelihood of choosing mechanism.

Again, we have a seemingly incongruous relationship between the Amount of Information score and mechanism; as in the overall score this may be a result of their not having the language to explain their ideas on the epistemology assessment more than their lack of conceptual understanding.

Discussion

These results highlight the highly contextual nature of epistemology. The two articles prompted students to frame the same task very differently, as they were designed to do by the journalists who wrote them, which resulted in the students applying epistemological resources in different ways that were not always correlated with how they had applied them to the epistemology assessment.

For Study 1, we kept our coding very broad, looking simply at whether students showed evidence of activating the very general ERs of Evidence, Cause and Effect, Correctness, and Amount of Information in their answers. This was because our focus was on establishing that epistemological change over two years can be captured by this type of performance assessment, rather than on a more detailed and nuanced description of the way students framed each task.

It is clear from these results that this coding was insufficient in order to use these assessments to predict how students would approach a task more grounded in reality, where most people's consumption of science is filtered through a media lens. Two original coding categories, Evidence and Correctness, did not predict anything. Cause and Effect was moderately useful, Causal Coherence predicted exactly one thing, and while Amount of Information was a significant predictor in two cases, both times the nature of the relationship was unclear and it seemed possible that it was really an indicator that the original category was a result of language limitations masking students' true thinking.

This doesn't mean that the assessment itself isn't potentially useful, though. After recoding Question 1 with more nuanced categories than simply presence or absence of some sort of attention to evidence, we found that the amount of attention students paid to the actual scientific data in that question as a valuable source of evidence significantly predicted their attention to the scientific research in these articles. It was also the only measure that predicted anything about their use of the anecdote in Article A, showing a negative association with the odds of a student both using the anecdote in their reasoning and underlining it as scientific evidence. Their responses to the anecdotal data in Question 1 also predicted some of their behavior on the news assessment, though that differed sharply according to which article they read—for those who read Article A it was associated with their attention to mechanism, while for those who read Article B it was associated with their attention to research.

This means that the last three out of four questions on the assessment were only occasionally useful in predicting student's behavior on the news assessment. It is possible that more detailed coding, paying more attention to students' value judgments about types of information rather than simply whether they thought about it or not, would yield better results. But even the one category that gave the best results from Question 1, Data Good Score, was a much weaker predictor of students' likelihood of believing that there was link between vaccines and autism and their attention to research than which article they read, and which article they read had a major impact on which scores predicted anything at all.

Implications for instruction

In general, this may speak to the fact that the epistemology assessment was designed with an eye to the IQWST science curriculum used in the schools we studied, and what epistemological development it is designed to promote, to give us the best chance of finding any epistemological changes that might exist. Some of what we are seeing may be an indication that the IQWST curriculum, while it focuses on engaging students in scientific practices in many ways that are productive, might not address the particular challenges that come with interpreting science as it is reported in the media.

For example, when discussing the particulate nature of matter, students observe what happens when they cover the end of a syringe and try to expand or contract the air inside. While these observations are a useful way to get students to think about the physical nature of air, and because the entire class does them there is the implicit acknowledgement that they can be replicated, they are not asked to evaluate the value of their own individual observations versus aggregate data from the entire class.

If we truly want to prepare students to engage with science in the real world after school, that means preparing them to engage with science as reported in the news media. That may be a very different task than preparing them to evaluate scientific explanations and arguments devoid of the media context. A news story about a socioscientific issue may not activate the same epistemological frames that a non-media task would, as might two stories written to deliberately frame the issue in two different ways, which may result in students applying very different sets of ERs in very different ways to each situation even if they all involve the same basic task of evaluating a scientific argument. It may be worthwhile to give students experience not only in engaging with SSIs, but specifically in engaging with them via authentic media accounts in all their complexity.

Implications for assessment

While our assessment, as written, may benefit from more nuanced coding in order to make it more useful in predicting students' behavior in other contexts, this study has also pointed out some inherent deficiencies in its design.

In Study 1, we mentioned that one problem was that students don't tend to talk about things whose absence isn't highlighted for them in some way. They could easily have brought up the lack of plausible mechanism as a deficiency of either argument in Question 1, but not a single student did. But in Question 2, when some explanations were designed to highlight deficiencies in other explanations, some students did attend to those. The only students who pointed out a lack of evidence in Question 2 were those who claimed that one explanation did have evidence and the others didn't (none of them did), none of whom elaborated beyond "it has/doesn't have evidence," which resulted in our not being able to interpret what they were actually referring to as "evidence." And no students mentioned a lack of evidence on Question 4. This resulted in many more chances to talk about cause and effect than about evidence.

Given that fact and the results of this analysis, we would likely replace one of the last three questions with a second question addressing evidence. Since Question 1 addresses a contrast between anecdotal evidence and scientific data, this would be an opportunity to present a similar contrast between mechanism and evidence. We might also redesign Question 1 to take out the cherry-picking of data issue; very few students picked up on this overall, and because of its high correlation with talking about the data it didn't prove useful to our analysis. This would provide a clearer distinction between the two types of evidence, and might make measures of whether students judge the anecdote positively or negatively more meaningful. Given that anecdotes are often included in news stories, as they are in Article A, because of their emotional impact, it may also be useful to give the anecdote used more emotional resonance than simply stating that one man felt better than he would have if he hadn't taken the headache medicine.

The fact that the Amount of Information Score indicated that students were *more* likely to attend to research and mechanism rather than less likely indicates that it is likely not measuring what we thought it was measuring. We acknowledged the possibility in Study 1 that students, particularly sixth graders, may not have the language to discuss all of the ideas that they implicitly understand. One student who says an explanation contains "more details" than another may simply mean that it includes more facts, while another student may say the same thing but mean that it contains more causal information, and it may not be possible to tease this apart without followup interviews. But the fact that both of the measures it predicted involved the underlining task suggests that these students may be more likely to identify particular information when asked to, even if they can't name it on their own. It might be useful, for example, in Question 2 to have them not only circle the best answer and explain why it's the best but also underline any specific sentences or phrases that make it better than the other answers. If students are truly relying solely on amount of information, they would be just as likely to underline an extraneous sentence as a sentence dealing with causal mechanism.

Future directions

This study raises many new questions to be addressed in future research. The simplest of these is whether there is any more information that the epistemology assessments, as written, can give us about students' approach to the news assessments, which could be addressed through recoding Questions 2 and 4 as we did Question 1. Beyond that, the above changes to a rewritten epistemology assessment may result in a tasks that students are more likely to frame in the same way they would an article like Article A, and thus give better information about their susceptibility to journalistic choices that may obscure scientific consensus.

More generally, more research is needed on students' reasoning about actual media stories about science, rather than about constructed texts that simulate some but not all features of a news article written entirely to meet a journalist's, rather than a researcher's, goals. We also need to pay more attention to whether our science education is actually preparing students to engage with science in this particular context, and whether our assessments are measuring that goal. We may even want to consider incorporating authentic media reports into assessments, rather than trying to determine whether the results of an assessment will hold in another context.

Chapter 5:

Synthesis and Discussion

The central question of this dissertation was: *How can we assess epistemological change in a way that gives us meaningful information about how people will reason about SSIs in the real world?*

Taken together, these three studies are an important start to answering this question. First, I found that we can measure epistemological change from an epistemological resources perspective with written performance tasks that engage students in scientific practices. Then I showed that journalists' choices have a strong impact on how students reason about a given SSI, including what information they base their decisions on and how they categorize that information. And finally, I found that while some measures from the first study allowed us to predict aspects of students' performance on the second, the differences between the articles overshadowed those predictions.

These findings have many implications, both for how we assess epistemological beliefs about science and how we should think about preparing students to engage with science in the real world.

Assessment of Epistemological Beliefs

I started out my assessment design with a list of six design principles. In light of the results of these studies, I would reconsider some of those in the following ways:

1. The situations provided need to resemble situations the learner may encounter in the real world and the tasks must ask learners to engage in practices that would be relevant to that situation.

We need to think about what "situations the learner may encounter in the real world" really means. On the one hand, writing tasks to resemble media accounts of science may provide a more realistic context; on the other hand, it would make it harder to piece apart their beliefs about science from the impact of journalism on those beliefs. In the end, does this distinction matter, if what we're truly concerned with is predicting behavior outside of school? It will take much more research to disentangle these issues. We need to look at other real-world situations beyond reading about SSIs in the news, other epistemological resources and uses of those resources that might be relevant, and how to measure media literacy skills and their interaction with science literacy.

3. Learners need to be able to express their thinking freely. If the questions are too tightly constrained, we may not get a complete enough picture of their thinking.

4. When possible, learners should be given multiple opportunities to express their thinking about each situation. They may not talk about every aspect of the situation that they are considering in a single answer.

These both must be balanced with the length of the assessment. Many students, particularly sixth graders, were not able to complete the entire epistemology assessment, and their data for Question 4 had to be dropped. These principles are vital to get a full picture of students' thinking, but are rendered useless if a student doesn't even get to the question. Future research can help us pinpoint exactly what resources and uses of those resources give us the most useful data about behaviors in other settings, so that we can design items that address those in the most economical way possible.

For example, Question 1 was designed to give students a chance to notice that while Peter was using the research data, he was using it selectively in a way that contradicted the overall trends of the data. While quite a few students did talk about this, it turned out not to be a useful predictor of their behavior on the news assessment; in fact, it was highly correlated enough with their judgment of the research data as a good source of evidence overall (a much better predictor) that it was dropped from the analysis entirely. It's possible that redesigning the question without this aspect would have allowed students to answer it more quickly without depriving us of any useful predictors.

6. The situations provided should give learners the opportunity to make use of the ERs that we believe would be productive when faced with a similar situation in real life. For example, not every instance of engaging productively in a scientific practice requires the learner to think about evidence, so we need to be sure we include at least one opportunity for them to do so if that is something we are interested in.

A corollary to this is that if we want students to notice the lack of something, there has to be a contrasting case to highlight that lack. In Question 2, this was done successfully—some explanations contained causal chains and some didn't, and many students noticed this lack. However, no students noticed the lack of plausible mechanisms in the arguments in Question 1, where neither argument contained one. And again, while these studies are a good start, we need more research to decide exactly which ERs are most useful to assess.

Implications for Education

This research points to both the complications and possible benefits of using media reports of science in the classroom.

In order to prepare students to reason about science that has been filtered through a journalistic lens, we may need to give them direct experience with it, and practice reasoning about it. In some schools where media literacy is addressed in other classes, it may be beneficial to design cross-disciplinary projects that develop both media literacy and science literacy simultaneously. In others, it may be necessary to introduce some level of media literacy instruction to the science curriculum. Students need to understand how to recognize biased writing (even if that bias is caused by a "balance" between two sides that contradicts the scientific consensus), how and why journalists appeal to emotions over reason, and learn to judge the different justifications and evidence presented.

For the same reasons, introducing news articles to the classroom without discussion of these issues may in some cases be counterproductive. Students may latch on to pieces of information that teachers do not see as convincing, coming to a different conclusion than intended. This is not a reason to avoid using news articles, but an indication that teachers must be willing to discuss and, if necessary, actively counteract any miscommunication that might be caused by the journalists' choices.

Future Directions

One thing that these studies indicate for certain is the need for additional attention to the interplay between media literacy and science literacy. As I noted in the introduction, most students we are educating will not be scientists; whether it is productive to engage such students in scientific practices at all, and if so whether and how scientific practice in the classroom for educational purposes should differ from the practices of actual scientists, has long been a topic of debate and discussion within the science education community. As more attention is paid to nonscientists' actual practices when engaging with science, we must continue to adjust our educational practices accordingly.

While an epistemological understanding of science gained through authentic practice is an important factor, there are other beliefs and skills that may not strictly be considered "scientific" but will have as much of or more of an impact on people's engagement with science. People have epistemological beliefs about media as well; for example, how it functions to filter and frame knowledge or the goals of the people who present knowledge through the media. It is vital that we do not ignore the development and application of these beliefs simply because they do not fall under the typical purview of "science." It is entirely possible that we *cannot* meet our goal of preparing students to engage with science-related issues in the real world without preparing them for the fact that they will be encountering these issues via the news media.

References

Albe, V. (2008) Students' positions and considerations of scientific evidence about a controversial socioscientific issue. *Science & Education*, *17*, 805-827

Albe, V. (2008b). When scientific knowledge, daily life experience, epistemological and social considerations intersect: Students' argumentation in group discussions on a socio-scientific issue. *Research in Science Education, 38*, 67–90.

Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education, 95*, 518-542.

Alvermann, D.E. (2008). Why bother theorizing adolescents' online literacies for classroom practice and research? *Journal of Adolescent & Adult Literacy*, *52*(1), 8-19.

Annotated bibliography

Arke, E.T., & Primack, B.A. (2011). *Quantifying new media literacy: Development, reliability, and validity of a new measure.* EMI Educational Media International, *46*(1), 53-65.

Bell, R. L., Blair, L. M., Crawford, B. A., & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching*, *40*(5), 487-509.

Berland, L. K., & Hammer, D. (2012a). Framing for scientific argumentation. *Journal of Research in Science Teaching*, *49*(1), 68–94.

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

Berland, L., & Crucet, K. (2016). Epistemological Trade-Offs: Accounting for Context When Evaluating Epistemological Sophistication of Student Engagement in Scientific Practices. *Science Education*, *100*(1), 5-29.

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

Boykoff, M.T., and Boykoff, J.M. (2004). Balance as bias: Global warming and the US prestige press. *Global Environmental Change, 14*, 125-136.

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*(4), 639–669. doi: 10.1002/sce.20449

Britt, M.A., Richter, T., & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist, 49*(2), 104-122.

Bromme, R., and Goldman, S.R. (2014). The public's bounded understanding of science. *Educational Psychologist*, *49*(2), 59-69.

Brown, J.D. (2006). Media literacy has potential to improve adolescents' health. *Journal of Adolescent Health*, *39*, 459-460.

Bubela, T., et al. (2009). Science communication reconsidered. *Nature*, *27*(6), 514-518.

Carvalho, A. (2007). Ideological cultures and media discourse on scientific knowledge: re-reading news on climate change. *Public Understanding of Science*, *16*, 223-243.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175–218.

Christodoulou, A., & Osborne, J. F. (2014). The science classroom as a site of epistemic talk: A case study of a teachers' attempts to teach science based on argument. *Journal of Research in Science Teaching,* 51(10), 1275–1300.

Clarke, C.E. (2008). A question of balance: The autism-vaccine controversy in the British and American elite press. *Science Communication,* 30(1), 77-107.

Dahlstrom, M. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences,* 111, 13614-13620.

Davis, E. (2003). Untangling dimensions of middle school students' beliefs about scientific knowledge and science learning. *International Journal of Science Education*, *25*(4), 439-468.

Dearing, J.W. (1995). Newspaper coverage of maverick science: Creating controversy through balancing. *Public Understanding of Science*, 4, 341-361.

Deephouse, D. L., & Suchman, M. (2008). Legitimacy in organizational institutionalism. *The Sage handbook of organizational institutionalism*, 49, 77.

Deng, F., Chen, D. T., Tsai, C. C., & Chai, C. S. (2011). Students' views of the nature of science: A critical review of research. *Science Education*, *95*(6), 961-999.

Dixon, G.N. and Clarke, C.E. (2012). Heightening uncertainty around certain science: Media coverage, false balance, and the autism-vaccine controversy. *Science Communication*, 35(3), 358-382.

Duggan, S., and Gott, R. (2002). What sort of science education do we really need? *International Journal of Science Education,* 24(7), 661-679.

Dunwoody, S. (2014). Science journalism: Prospects in the digital age. In Bucchi, M. and Trench, B. (Eds.), *Routledge Handbook of Public Communication of Science and Technology.*

Duschl, R. A. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32(1), 268–291.

Eggert, S. & Bögeholz, S. (2009). Students' use of decision-making strategies with regard to socioscientific issues: An application of the Rasch partial credit model. *Science Education*, 94, 230-258.

Elby, A. (2000). What students' learning of representations tells us about constructivism. *The Journal of Mathematical Behavior*, *19*(4), 481-502.

Elby, A. (2009). Defining personal epistemology: A response to Hofer & Pintrich (1997) and Sandoval (2005). *The Journal of the Learning Sciences*, *18*(1), 138-149.

Elby, A., & Hammer, D. (2010). Epistemological resources and framing: A cognitive framework for helping teachers interpret and respond to their students' epistemologies. *Personal epistemology in the classroom: Theory, research, and implications for practice*, 409-434.

Guzzetti, B. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade in research? *Reading & Writing Quarterly*, 16, 89-98.

Halkia, K., and Mantzouridis, D. (2005). Students' views and attitudes towards the communication code used in press articles about science. *International Journal of Science Education*, 27(12), 1395-1411.

Hammer, D. & Elby, A. (2002). On the form of a personal epistemology. In Hofer, B.K., & Pintrich, P.R. (Eds.), *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing,* (pp. 169-190). Mahwah, NJ: Erlbaum.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. *Transfer of learning from a modern multidisciplinary perspective*, *89*.

Hobbs, R. (1998). The seven great debates in the media literacy movement. *Journal of communication*, 48(1), 16-32.

Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Educational Psychology Review*, *13*(4), 353-383.

Hutchison, P., & Hammer, D. (2010). Attending to student epistemological framing in a science classroom. *Science Education*, *94*(3), 506-524.

Jensen, J. D., & Hurley, R. J. (2012). Conflicting stories about public scien- tific controversies: Effects of news convergence and divergence on sci- entists' credibility. *Public Understanding of Science*, 21, 689–704. doi:10.1177/0963662510387759

Jensen, J.D. (2008). Scientific uncertainty in news coverage of cancer research: Effects of hedging on scientists' and journalists' credibility. *Human Communication Research*, 34, 347-369.

Jiménez-Aleixandre, M.P. and Federico-Agraso, M. (2009). Justification and persuasion about cloning: Arguments in Hwang's paper and journalistic reported versions. *Research in Science Education*, 39, 331-347.

Jiménez-Aleixandre, M.P., and Puig, B. (2012). Argumentation, evidence evaluation, and critical thinking. In Frazer, B.J. et al. (Eds.), *Second International Handbook of Science Education*, Springer.

Karisan, D. & Zeidler, D.L. (2017). Contextualization of nature of science within the socioscientific issues framework: A review of research. *International Journal of Education in Mathematics, Science and Technology*, 5(2), 139-152.

Kelly, G.J. and Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Educa*tion, 86, 314-342.

Khishfe, R. (2012). Relationship between nature of science understandings and argumentation skills: A role for counterargument and contextual factors. *Journal of Research in Science Teaching*, 49(4), 489-514.

Kolstø, S.D. (2001). 'To trust or not to trust...' - Pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education*, 23(9), 877-901.

Kolstø, S.D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education*, 85, 291-310.

Kolstø, S.D., et al. (2006). Science students' critical examination of scientific information related to socioscientific issues. *Science Education*, 90, 632-655.

Korpan, C.A., Bisanz, G.L., Bisanz, J., and Henderson, J.M. (1997). Assessing literacy in science: Evaluation of scientific news briefs. *Science Education*, 81, 515-532.

Krajcik, J., Reiser, B.J., Sutherland, L.M., & Fortus, D. (2004). *IQWST: Investigating and questioning our world through science and technology*. Ann Arbor, MI: University of Michigan.

Lederman, N.G., Abd El-Khalick, F., Bell, R.L, & Schwartz, R.S. (2002). Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science. *Journal of Resesarch in Science Teaching, 39(6),* 497- 521.

Lee, Y.C., and Grace, M. (2012). Students' reasoning and decision making about a socioscientific issue: A cross-context comparison. *Science Education*, 96, 787-807.

Leung, J.S.C., Wong, A.S.L., and Yung, B.H.W. (2014). Understandings of nature of science and multiple perspective evaluation of science news by non-science majors. *Science & Education*, 24, 887-912.

Liu, S.-Y., Lin, C.-S., and Tsai, C.-C. (2011). College students' scientific epistemological views and thinking patterns in socioscientific decision making. Science Education, 95, 497-517.

Logan, R.A., Zengjun, P., and Wilson, N.F. (2000). Prevailing impressions in science and medical news: A content analysis of the Los Angeles Times and The Washington Post. *Science Communication*, 22(1), 27-45.

Louca, L., Elby, A., Hammer, D., & Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. *Educational Psychologist*, *39*(1), 57-68.

Maier, M., Rothmund, T., Retzbach, A., Otto, L., & Besley, J.C. (2014). Informal learning through science media usage. *Educational Psychologist*, 49(2), 86-103.

McClune, B., & Jarman, R. (2010). Critical reading of science-based news reports: Establishing a knowledge, skills, and attitudes framework. *International Journal of Science Education*, 32(6), 727-752.

Mikulak, A. (2011). Mismatches between 'scientific' and 'non-scientific' ways of knowing and their contribution to public understanding of science. *Integrative Psychological and Behavioral Science,* 45, 201-215.

Moyer-Gus e, E., & Nabi, R. L. (2010). Explaining the effects of narrative in an entertainment television program: Overcoming resistance to per- suasion. *Human Communication Research*, 36, 26–52. doi:10.1111/ j.1468-2958.2009.01367.x

National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School.* Washington, D.C.: National Academies Press.

National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press. doi:https://doi.org/10.17226/13165.

National Science Board. 2016. *Science and Engineering Indicators 2016*. Arlington, VA: National Science Foundation (NSB-2016-1). Accessed at https://www.nsf.gov/statistics/2016/nsb20161/#/

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.

Osborne, J.F., et al. (2016). The development and validation of a learning progression for argumentation in science. *Journal of Research in Science Teaching*, 53(6), 821-846.

Polman, J.L., Newman, A., Saul, E.W., and Farrar, C. (2014). Adapting practices of science journalism to foster science literacy. *Science Education*, 98, 766-791.

Potter, W.J. (2004). Argument for the need for a cognitive theory of media literacy. *The American Behavioral Scientists*, 48(2), 266-272.

Ranger, M., & Bultitude, K. (2016). 'The kind of mildly curious sort of science interested person like me': Science bloggers' practices relation to audience recruitment. *Public Understanding of Science,* 25(3), 361-378.

Ratcliffe, M. (1999). Evaluation of abilities in interpreting media reports of scientific research. *International Journal of Science Education*, 21(10), 1085-1099.

Raveendran, A., and Chunawala, S. (2015). Values in science: Making sense of biology doctoral students' critical examination of a deterministic claim in a media article. *Science Education*, 99, 669-695.

Reis, P., & Galvão, C. (2004) Socio-scientific controversies and students' conceptions about scientists, *International Journal of Science Education*, 26:13, 1621-1633

Rivadeneyra, R., and Ward, L.M. (2005). From Ally McBeal to Sábado Gigante: Contributions of television viewing to the gender role attitudes of Latino adolescents. *Journal of Adolescent Research*, 20(4), 453-475.

Rosenberg, Hammer, & Phelan 2006 Multiple epistemological coherences in an eighth grade discussion of the rock cycle. *The Journal of the Learning Sciences,* 15, 261

Ruhrmann, G., Guenther, L., Kessler, S.H., and Milde, J. (2015). Frames of scientific evidence: How journalists represent the (un)certainty of molecular medicine in science television programs. *Public Understanding of Science*, 24(6), 681-696.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education*, *93*(5), 875-891.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499-525.

Sadler, T. D., & Zeidler, D. L. (2004). The morality of SSI: Construal and resolution of genetic engineering dilemmas. *Science Education, 88,* 4–27.

Sadler, T.D. (2004). Informal reasoning regarding socioscientific issues: A critical review of the research. *Journal of Research in Science Teaching, 41*(5), 513-536.

Sadler, T.D., Barab, S.A., and Scott, B. (2007). What do students gain by engaging in socioscientific inquiry? *Research in Science Education, 37*, 371-391.

Sadler, T.D., Chambers, F.W., and Zeidler, D.A. (2004). Student conceptualizations of the nature of science in response to a socioscientific issue. *International Journal of Science Education, 26*(4), 387-409.

Salter, I. Y., & Atkins, L. J. (2014). What students say versus what they do regarding scientific inquiry. *Science Education*, *98*(1), 1-35.

Sampson, V., Grooms, J., and Walker, J.P. (2010). Argument-driven inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education, 95*, 217-257.

Sandoval, W.A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12(1),* 5-51.

Sandoval, W.A., & Cam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education, 95*, 383-408.

Sandoval, W.A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist, 49*(2), 139-152.

Schäfer MS (2012) Taking stock: A meta-analysis of studies on the media's coverage of science. *Public Understanding of Science 21*(6): 650–663.

Scherr, R.E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics, *Cognition and Instruction, 27*(2), 147-174, DOI: 10.1080/07370000902797379

Schommer-Aikins, M., and Hutter, R. (2002). Epistemological beliefs and thinking about everyday controversial issues. *The Journal of Psychology, 136*(1), 5-20.

Seakins, A., & Hobson, M. (2017). Public understanding of science. In Taber, K.S. & Akpan, B. (Eds.) *Science Education: An International Course Companion*. Rotterdam: SensePublishers.

Sinatra, G.M., Kienhues, D., & Hofer B.K. (2014) Addressing Challenges to Public Understanding of Science: Epistemic Cognition, Motivated Reasoning, and Conceptual Change, *Educational Psychologist, 49*(2), 123-138, DOI: 10.1080/00461520.2014.916216

Smith & Wenk 2006

Sørvik, G.O., Blikstad-Balas, M., and Ødegaard, M. (2015). "Do books like these have authors?" New roldes for text and new demands on students in integrated science-literacy instruction. *Science Education, 99*, 39-69.

Stocking, S. H., & Holstein, L. W. (2009). Manufacturing doubt: journalists' roles and the construction of ignorance in a scientific controversy. *Public Understanding of Science*, *18*(1), 23-42.

Strømsø, H.I., and Bråten, I. (2010). The role of personal epistemology in the self-regulation of internet-based learning. *Metacognition and Learning, 5,* 91-111.

Tal, T., & Kedmi, Y. (2006). Teaching socioscientific issues: Classroom culture and students' performances. *Cultural Studies of Science Education*, *1*(4), 615-644.

Thomm, E., and Bromme, R. (2011). "It should at least seem scientific!" Textual features of "scientificness" and their impact on lay assessments of online information. *Science Education, 96*, 187-211.

Topcu M.S., Sadler, T.D., & Yilmaz-Tuzun, O. (2010) Preservice Science Teachers' Informal Reasoning about Socioscientific Issues: The influence of issue context, *International Journal of Science Education, 32*(18), 2475-2495

Tseng, Y.-H., Chang, C.-Y., Rundgren, S.-N. C., and Rundgren, C.-J. (2010). Mining concept maps from news stories for measuring civic scientific literacy in media. *Computers & Education, 55*, 165-177.

Tytler, R., Duggan, S., and Gott, R. (2001). Dimensions of evidence, the public understanding of science, and science education. *International Journal of Science Education, 23*(8), 815-832.

Weaver EL and Bimber B (2009) Searching for a frame: News media tell the story of technological progress, risk, and regulation. *Science Communication 31*, 139–166.

White, D. M. (1950). The gate keeper: A case study in the selection of news. *Journalism Quarterly, 27,* 383–390.

Wickman 2004 The practical epistemologies of the classroom: A study of laboratory work. *Science Education, 88*(3), 325

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model- based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*(5), 941-967.

Wong, J. and Sandoval, W.A. (2016, April 11). *Effects of goal priming on high school students' use of mechanism and evidence information in a science media text*. Paper presented at 2016 annual meeting of the American Educational Research Association.

Wu, Y.-T., & Tsai, C.-C. (2011) High School Students' Informal Reasoning Regarding a Socio-scientific Issue, with Relation to Scientific Epistemological Beliefs and Cognitive Structures, *International Journal of Science Education, 33*(3), 371-400,

Yaros, R.A. (2011). Effects of text and hypertext structures on user interest and understanding of science and technology. *Science Communication, 33*(3), 275-308.

Zeidler, D. L., Herman, B. C., Ruzek, M., Linder, A., & Lin, S. S. (2013). Cross-cultural epistemological orientations to socioscientific issues. *Journal of Research in Science Teaching*, *50*(3), 251-283.

Zeidler, D. L., Walker, K. A., Ackett, W. A., & Simmons, M. L. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science education*, *86*(3), 343-367.

Zeidler, D.L., Herman, B.C., Ruzek, M., Linder, A., and Lin, S.-S. (2013). Cross-cultural epistemological orientations to socioscientific issues. *Journal of Research in Science Teaching, 50*(3), 251-283.

Zeidler, D.L., Sadler, T.D., Applebaum, S., and Callahan, B.E. (2009). Advancing reflective judgment through socioscientific issues. *Journal of Research in Science Teaching, 46*(1), 74-101.

Zeidler, D.L., Walker, K.A., Ackett, W.A., and Simmons, M.L. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science Education, 86,* 343-367.

Appendices

Appendix A: Epistemology Assessment

Name: ___________________________________

Teacher: ________________________ Grade Level: ____

Male / Female

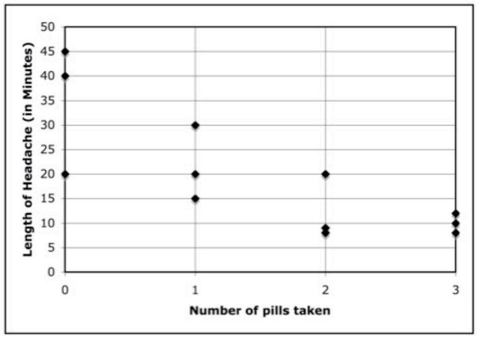
Date:____________________

Thank you for taking this questionnaire. There are four pages of questions after this one. Please read every question carefully and be as specific and complete in your answers as possible. If you need extra space to answer any of the questions, please use the back of the page.

There are no right or wrong answers to these questions. Your science teacher will not see these questionnaires, and it will not impact your grade in science class in any way.

1. A company has created a new pill to treat headaches. They had twelve people take the pill when they got a headache, and measured how long it took for the headache to go away. Different people were given different numbers of pills. The results can be seen in the table and graph below:

Number of Pills Taken	Length of Headache (in Minutes)
0	45
0	20
0	40
1	30
1	20
1	15
2	8
2	20
2	9
3	10
3	8
3	12



Two employees of the company, Anne and Peter, disagree on whether the pill works or not. Here is their conversation:

Anne: This pill is clearly a success.
Peter: No, our pill is a failure and we need to make a new one. Look, this person who took two pills had a headache 20 minutes long, and someone who took no pills also had a headache that was 20 minutes long! If taking two pills doesn't make your headache go away faster than not taking pills, the pills don't work.
Anne: That's a coincidence. Overall, our pill definitely works.
Peter: Also, two of the people who took three pills had longer headaches than two of the people who took two pills. The pill does not work.
Anne: I talked to one of the people who took one pill, and he said he definitely felt better than he would have without taking it.
Peter: He doesn't know that for sure. One person who took one pill had a longer headache than a person who took no pills. Maybe he would have had a shorter headache if he hadn't taken it.

A. Do you think that the pills work? Why or why not?

____

____

____

____

B.What are the strengths and weaknesses of each argument?

	Anne	Peter
Strengths		
Weaknesses		

C. Whether or not their answers are right or wrong, who do you think did a better job of arguing for their answer, Anne or Peter? Why?

____

____

____

____

____

2. Mrs. Arnold keeps a fishtank in her classroom that is set to 70 degrees. One weekend, the janitor accidentally ran into the fishtank and bumped the temperature setting to 90 degrees. Some of Mrs Arnold's fish had started to die by the time she got back to the classroom on Monday.

Here are some possible explanations for what happened:

A. 90 degrees is too hot for fish to live.

B. To survive, fish take in water through their gills and breathe the dissolved oxygen in the water. Warmer water holds less dissolved oxygen than cooler water. When the temperature of the water was turned up, it couldn't hold enough oxygen for all of the fish in the tank. Some fish started dying because they did not have enough oxygen and couldn't breathe.

C. Hotter water holds less dissolved oxygen than cooler water, so it killed some fish.

D. To survive, fish need to eat food and take in water through their gills to breathe the dissolved oxygen in it. Warmer water holds less dissolved oxygen than cooler water. When the temperature of the water was turned up, it held less oxygen and felt hotter to the fish. Some fish started dying because they did not have enough oxygen, and the ones that were left were probably uncomfortable and hungry.

E. Fish can only live in water that is a certain temperature. If you turn the water down too cold or turn it up too hot, fish cannot survive in the water. In this case, the water was turned up too hot when it was turned up an additional 20 degrees. Some fish died because the janitor ran into the tank and turned the temperature up to 90 degrees.

Which is the *best scientific explanation* for why some of the fish died? Circle the answer that is the best scientific explanation.
For the answer you circled, write why you think it is the best scientific explanation. For all of the other answers, write why you don't think that it is as good as the best explanation. Be as *specific* as you can in your answers - for example, instead of writing "This answer has more details," write which details are important and why they are important.

A. ____

____

____

B. ____

____

____

C. ____

____

____

D. ____

____

____

____

E. ____

____

____

____

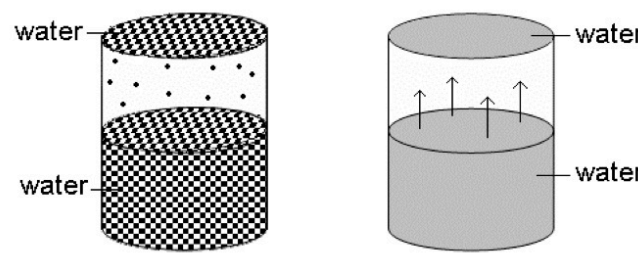
3. Choose ONE of these questions and write *a scientific explanation* to answer the question:

a. When a person is wearing perfume across the room, how can you smell the perfume?
b. If you leave a dish of water sitting out overnight, why will the water be gone the next day?
c. An area contains only herbivores and plants. A pack of carnivores moves into the area. A few months after the carnivores have moved in, why would the plant population go up?

4. Imagine that you and all of your classmates are asked to assess each other's projects in the science fair. There are two projects that present models of evaporation and condensation.

Cosette's model: Cosette has drawn water in a jar as lots and lots of dots, water in the air as some dots, and water on the plastic wrap covering the jar as lots of dots.

Makayla's model: Makayla has drawn blue water in a jar, arrows to show that water is moving to the plastic wrap, and blue water on the plastic wrap.



Cosette's model *Makayla's model*

A. What would you tell Cosette and Makayla about the strengths and weaknesses of each of their models? Write down your answers in the table below.

	Cosette's Model	Makayla's Model
Strengths		
Weaknesses		

B. Which is a better scientific model for evaporation and condensation? Why?

C. How could you revise Cosette's model to make it better?

D. Why would you make those changes?

E. How would you revise Makayla's model to make it better?

F. Why would you make those changes?

G. Could you use any of the ideas in Cosette's or Makayla's models to explain how we can smell perfume from a distance? If so, how? If not, why not?

Appendix B: News Assessment Article A

FORM A

Name: ________________________________

Teacher: ______________________ Grade Level: ____

Male / Female

Date:__________________

Thank you for taking this questionnaire. After this page, there is a two page news article followed by three pages of questions. Please read every question carefully and be as *specific* and *complete* in your answers as possible. If you need extra space to answer any of the questions, please use the back of the page.

There are no right or wrong answers to these questions. Your science teacher will not see these questionnaires, and it will not impact your grade in science class in any way.

Please read the following news article carefully. This article is from a website that reports national news, such as CNN.com. It is okay if there are some parts of the article you don't understand, since it is a news article written for adults. Just do your best to decide what you think it means. You may refer back to the article as necessary to answer the questions afterward.

Vaccine-autism question divides parents, scientists

YUMA, Arizona -- At 13, Michelle Cedillo can't speak, wears a diaper and requires round-the-clock monitoring in case she has a seizure. While her peers go to school or the mall or spend time with friends, the Yuma, Arizona, teenager remains at home, where she entertains herself with picture books and "Sesame Street" and "Blue's Clues" DVDs.

Michelle has no idea she is at the center of a court case pitting thousands of families of children with autism against the medical establishment. A number of prestigious medical institutions say there is no link between vaccines and autism. The families believe vaccines caused their children's autism, and they've taken their case to court.

"I think there is a link," says Theresa Cedillo, Michelle's mother.

Theresa and her husband, Mike, say their only child was a happy, engaged toddler who responded to her name, said "mommy" and "daddy" and was otherwise normal until she received a measles, mumps and rubella (MMR) vaccine at 15 months.

They believe the MMR vaccine, which contained mercury, drastically altered the course of their daughter's development. Within days of receiving the injection, Michelle suffered from a high fever, persistent vomiting and problems with her digestion. Worse still, her parents say, Michelle stopped speaking and no longer responded to her name.

"I thought it was because she was so sick. I thought certainly she'll start talking again," Theresa recalls. Michelle has since been diagnosed with autism, inflammatory bowel disease, arthritis, osteoporosis and epilepsy.

Dr. Paul Offit, chief of infectious diseases at the Children's Hospital of Philadelphia, Pennsylvania, says the connection between vaccines and autism is nothing more than a sad coincidence.

"About 20 percent of children with autism will regress between their first and second birthday," says Offit. "So statistically, it will have to happen where some children will get a vaccine. They will have been fine. They will get the vaccine, and they will not be fine anymore. And I think parents can reasonably ask the question, 'Is it the vaccine that did this?'"

The answer is no, according to the Centers for Disease Control and Prevention, the World Health Organization, and the Institute of Medicine.

In reaching its conclusion, the Institute of Medicine pointed to five large studies finding no link between autism and vaccines which contain mercury, and 14 large studies finding no link between the MMR vaccine and autism. Childhood vaccines no longer contain mercury, though it remains in some flu shots.

The studies compared autism rates among populations of children who did and did not receive the MMR vaccines, and among those who did and did not receive vaccines containing mercury.

"It's been asked and answered: Vaccines don't cause autism," Offit says.

Michelle Cedillo's parents disagree. They've sued the government through the National Vaccine Injury Compensation Program, established in 1988 to pay damages to those who have suffered as a result of vaccines.

Three attorneys appointed by the court are considering whether a combination of the MMR and the mercury in vaccines could cause autism; whether MMR vaccines alone could cause autism; and whether mercury alone could cause autism. To prove their case, families need to show a plausible biological mechanism for vaccines to cause autism.

At Michelle Cedillo's hearing last year, Dr. Marcel Kinsbourne, a pediatric neurologist who is a professor at The New School in New York, testified that he thought the measles vaccine was a "substantial factor" in causing the girl's autism. Traces of the measles virus were found in Michelle's gut, leading the Oxford University-trained doctor to conclude the girl's immune system had not rejected the virus. Kinsbourne told the court the measles virus invaded cells in Michelle's brain, resulting in her autism.

Already, families who believe vaccines can trigger autism are pointing to the case of 9-year-old Hannah Poling as a major victory. In November, the government conceded that vaccines "significantly aggravated" the Georgia girl's underlying illness, predisposing her to symptoms of autism.

"Vaccines save lives," countered Dr. Julie Gerberding, director of the Centers for Disease Control and Prevention, after the Poling decision was announced in March. "The most reputable scientists around the world have looked at this situation over and over again, and they have stated that they cannot see an association between vaccines and autism."

No matter what the vaccine court decides, the debate will most likely not go away anytime soon, nor will the sometimes devastating symptoms of the children with autism.

At the Cedillos' modest home, Michelle receives nutrition through a feeding tube, cannot walk unassisted and uses hand motions and tapping to communicate with her mother, father, grandmother and grandfather, who are her constant companions and caregivers.

Please refer back to the article as often as you need to to answer the following questions. There are no right or wrong answers; these questions are about what *you* think. Please give as much detail as possible when answering, so that you fully explain your thoughts. If you need more room to answer any questions, please continue on the back.

1. After reading this article, do you think that vaccines cause autism? Yes / No

Why or why not?

____

____

____

____

____

2. Did anything you read in the article affect your answer to Question 1? If so, what?

____

____

____

____

____

3. Do you think that there is more scientific evidence that vaccines do cause autism, or that they do not? Do / Do Not

Why?

4. Please **underline** any parts of the article that you think give scientific evidence in *either direction*. (Scientific evidence that vaccines either do cause autism, or that they don't.)

5a. The article says, "To prove their case, families need to show a plausible biological mechanism for vaccines to cause autism." What do you think they mean by "biological mechanism"?

5b. Why do you think it is important to show a "plausible biological mechanism"?

6. Is there anything in the article that you do not believe is true? If so, what, and why don't you believe it? Why do you think the author included it if it is not true? If there is nothing you do not believe, go to Question 7.

7. Do you think that anything in this article was included specifically to catch your attention or to affect your emotions? Please **circle** any parts that you think were included for these reasons. Why do you think they were included?

Appendix C: News Assessment Article B

FORM B

Name: ______________________________

Teacher: ____________________ Grade Level: ___

Male / Female

Date:_________________

Thank you for taking this questionnaire. After this page, there is a two page news article followed by three pages of questions. Please read every question carefully and be as *specific* and *complete* in your answers as possible. If you need extra space to answer any of the questions, please use the back of the page.

There are no right or wrong answers to these questions. Your science teacher will not see these questionnaires, and it will not impact your grade in science class in any way.

Please read the following news article carefully. This article is from a website that reports national news, such as CNN.com. It is okay if there are some parts of the article you don't understand, since it is a news article written for adults. Just do your best to decide what you think it means. You may refer back to the article as necessary to answer the questions afterward.

Autism's Link to Vaccines Is Unfounded, Study Says

Use of mercury in vaccines doesn't cause autism among school-aged children, a study says.

Advocacy groups had suggested that mercury used as a preservative in some vaccines may be linked to autism.

The Montreal study found that autism rates rose steadily among school-aged children who had been vaccinated as infants, even though the preservative was not included in the shots during the last two years examined, said Eric Fombonne, director of Pediatric Psychiatry at Montreal Children's Hospital

"If you look at the rates of autism in the study, there is a smooth increase in the prevalence from the group born in 1987 to when the study finishes with children born in 1998," said Fombonne, the lead researcher. Had mercury played a role, the rates of diagnoses should have dropped when mercury was removed from the vaccines in 1996, he said. The study, published in the journal Pediatrics, examined rates of the disorder among 27,749 children at 55 schools in Montreal.

Fombonne has worked as an expert witness on behalf of vaccine makers defending lawsuits in the U.S. that allege that mercury causes autism. While none of Fombonne's research is industry-funded, the study was immediately criticized as biased by David Ayoub, an Illinois radiologist.

`Heavily Biased'

"This is just another heavily biased study by an author with a long track record of financial ties to the drug industry, and whose previous views have been discredited," wrote Ayoub, who is also medical director of the Foundation for Autism Information and Research.

Ayoub said Fombonne's effort ignored several papers that link mercury and autism. These omissions and Fombonne's testimony on behalf of drugmakers raise issues about his ability to independently analyze the study data, Ayoub said.

The study shows a four-fold increase in autism rates during the 10-year period examined, Ayoub said. In addition, some children may have been exposed to mercury from other vaccines, including hepatitis and flu shots given to children and pregnant women after 1996 that still contained mercury, he said in an interview.

Fombonne said the increased autism rate is more likely tied to changes in the definition of autism, greater awareness of the condition, and rising numbers of services for patients. These have boosted the number of children diagnosed with the condition each year.

This isn't the first study to eliminate mercury as a possible cause for autism. The U.S. government-chartered Institute of Medicine in Washington issued a report in May 2004 concluding from a review of previous studies that no evidence linked autism with mercury-containing vaccines or with shots that no longer contain mercury.

The rates of disorders "among children born in an era where there was no exposure to mercury is significantly higher than in the years before, when there was high exposure," he said. "That's really a convincing argument that there is no relationship between the two."

Please refer back to the article as often as you need to to answer the following questions. There are no right or wrong answers; these questions are about what *you* think. Please give as much detail as possible when answering, so that you fully explain your thoughts. If you need more room to answer any questions, please continue on the back.

1. After reading this article, do you think that vaccines cause autism? Yes / No

Why or why not?

____

____

____

____

____

2. Did anything you read in the article affect your answer to Question 1? If so, what?

____

____

____

____

____

3. Do you think that there is more scientific evidence that vaccines do cause autism, or that they do not? Do / Do Not

Why?

4. Please **underline** any parts of the article that you think give scientific evidence in *either direction*. (Scientific evidence that vaccines either do cause autism, or that they don't.)

5a. The article quotes both Eric Fombonne and David Ayoub. Which of these two people do you trust more? Why?

5b. Why do you think the author of the article talked to both of these people?

6. Is there anything in the article that you do not believe is true? If so, what, and why don't you believe it? Why do you think the author included it if it is not true? If there is nothing you do not believe, go to Question 7.

____

____

____

____

____

____

7. Do you think that anything in this article was included specifically to catch your attention or to affect your emotions? Please **circle** any parts that you think were included for these reasons. Why do you think they were included?

____

____

____

____

____

____