ABSTRACT
The dominant practices in science are not discovery and justification but rather the broadening and deepening of explanations via theory development and conceptual modification. New tools, theories, and technologies are fundamentally changing the methods and inquiry practices of science. In turn, the forms of evidence and criteria for what counts as evidence change, too. The implication is the importance of including and understanding the epistemic and social practices embedded in the science learning goals ‘critique and communication of scientific ideas and information’. The study examines 11-year old students’ argumentation patterns and epistemic reasoning. The focus is on students’ arguments and choices as they pertain to reasons and evidence given to support claims about data and measurement. Results suggest that students are able to evaluate the quality of arguments made by classmates and can do so employing a number of different reasons or perspectives. An understanding of how students choose and use evidence is important for understanding how to coordinate formative assessments and focus teacher’s feedback on students’ epistemic reasoning.

KEYWORDS: Argumentation, Computer Supported Learning, Scientific Reasoning, Philosophy of Science.

Introduction
The last 50 years have been a parade of new scientific methods, ideas and beliefs; Pluto is no longer considered a planet! The practices and images of science today are so very different and this is largely due to the advancements in methods and instrumentation for observing, measuring, analyzing and modeling. Certainly the fundamental frameworks of scientists working from material evidence to propose explanations and mechanisms are generally the same. New tools, theories, and technologies though have fundamentally changed both the guiding frameworks and the inquiry practices of science (Duschl & Grandy, 2008; NRC, 2007a). In turn, the forms of evidence and criteria for what counts as evidence have had to change, too.

The contemporary understanding of the nature of science (c.f., Giere, 1988; Longino, 2002; Godfrey-Smith, 2003; Solomon, 2008) holds that the majority of scientists’ engagement is not individual efforts toward establishing final theory acceptance, but rather communities of scientists striving for improvement and refinement of a theory. What occurs in science is not predominantly the context of discovery or the context of justification but the contexts of theory development, of conceptual modification. Thagard (2007) posits that coherence, truth, and explanatory coherence of scientific explanations are achieved through the complementary process in which theories broaden and deepen over time by accounting for new facts and providing explanations of why the theory works. Recent research reviews (NRC, 2007a; Duschl, 2008; Duschl & Grandy, 2008; Ford & Forman, 2006; Lehrer & Schaub, 2006b) and research studies on science learning (Ford, 2008a; Lehrer, Schaub, & Lucas, 2008; Smith, Wiser, Anderson & Krajcik, 2006) maintain that the same broadening and deepening practices based on a set of improving and refining tenets within a community of investigators ought to hold for teaching and learning practices in science learning environments. For scholars in the emerging domains of the learning sciences and science studies the emphasis is on ‘science as practice’ and in particular the epistemic and social practices embedded in the critique and communication of scientific ideas and information.

The research reported here is part of a broader research program that seeks to better understand how to mediate students’ argumentation discourse (Duschl, 2008; Erduran & Jimenez-Alexandre, 2008; Duschl & Osborne, 2002). Here we report results of a study that examines students’ reasoning as it pertains to reasons and evidence given to support claims about data and measurement; i.e., what counts as evidence for the critique of data. An understanding of how students choose and use evidence is important to understanding how to coordinate and focus teacher’s feedback on epistemic reasoning and learners’ practices with building and refining models.

Review of Literature
Wellington and Osborne (2001) argue that if science and scientists are epistemically priviledged, then it is a major
shortcoming of our educational programs that we offer little to justify the current lack of focus on epistemic practices in classrooms. Wellington and Osborne are speaking to the misplaced priorities we find in most science curriculum. That is, the dominant focus is teaching what we know. How we came to know and why we believe what we know are marginalized aspects of science learning. Successful science education depends on students’ involvement in forms of communication and reasoning that models the discourse that occurs in scientific communities (e.g., Herrenkohl, Palincsar, DeWaters, Kawasaki, 1999; Roseberry, Warren & Conant, 1992; Schuble, Glaser, Duschl, Schulze, & John, 1995).

A critical step forward is engaging learners in examining the relationships between evidence and explanation. A dominant dynamic in scientific communication is the dialectical practices that occur between evidence and explanation or, more generally, between observation and theory. Increasingly when examined over the last 100 years, these dialectical exchanges involve or depend on the tools and instruments scientists employ (Zammito, 2004). Ackerman (1985) refers to such exchange practices as the ‘data texts’ of science and warns that the conversations among contemporary scientists about measurement, observations, data, evidence, models, and explanations is of a kind that is quite foreign from the conversations found in the general population and, we might add, in science classrooms (Duschl, 2008; Driver, Newton & Osborne, 2000).

Pickering (1995) referred to this conflation of conversations when describing experiments in high-energy physics as the ‘mangle of practice’. Hacking (1988) writes that it is the richness, complexity and variety of scientific life that has brought about a focus on ‘science as practice.’ For Pickering, scientific inquiry during its planning and implementation stages is a patchy and fragmented set of processes mobilized around resources. Planning is the contingent and creative designation of goals. Implementation for Pickering (1989) has three elements:

- a “material procedure” which involves setting up, running and monitoring an apparatus;
- an “instrumental model,” which conceives how the apparatus should function; and
- a “phenomenal model,” which “endows experimental findings within meaning and significance … a conceptual understanding of whatever aspect of the phenomenal world is under investigation.

The “hard work” of science comes in trying to make all these work together (Zammito, 2004; pp. 226-227).

A critical step forward toward changing the ‘what we know’ instructional condition is to engage learners in examining the relationships between evidence and explanation. As stated above, a dominant dynamic in scientific communication is the data text discourse that occurs between evidence and explanation. This dialectic typically takes the form of arguments and is central to science-in-the-making activities.

Duschl (2003; 2008) has developed a set of design principles that coordinate conceptual, epistemic and social learning goals. The pedagogical model is the Evidence-Explanation (E-E) continuum depicted schematically in Figure 1.

The E-E continuum (Duschl, 2003) has its roots in perspectives from science studies and connects to cognitive and psychological views of learning. The appeal to adopting the E-E continuum as a pedagogical framework for science education is that it helps work out the details of the epistemic discourse processes. It does so by formatting into the instructional sequence select junctures of reasoning, e.g., data texts transformations. At each of these junctures or transformations, instruction pauses to allow students to make and report judgments. The three critical transformations or judgments in the E-E continuum include:

a) Selecting or generating data to become evidence.
b) Using evidence to ascertain patterns of evidence and models.
c) Employing the models and patterns to propose explanations.

The results reported in this paper focus on the first and second transformations; i.e., the reasons given for what counts as evidence and patterns in evidence. Yet another important judgment is deciding what data to obtain (Lehrer & Schauble, 2006a; Petrosino, Lehrer & Schauble, 2003). The development of children’s consideration for measurement and data prior to starting the E-E continuum is critically important, too.

Research Context and Methods

The present study examines two classes of 11-year old students’ argumentation patterns and epistemic reasoning. The school setting was a K-12 international school for American children located in a major metropolitan city. One of many located around the world, the school provides an American education for families living abroad. The school is fee paying and the students come from a mix of middle and high SES families. The school was physically divided into three sections...
general stages: facilitating and mediating discourse practices and it has three:

1. Receive ideas. Students ‘show what they know’ by producing detailed writing, drawings, and storyboards.
2. Recognize ideas. The diversity of ideas received is made public. Students’ ideas and the data from investigations are then discussed against a set of scientific criteria and evidence. The teacher selects the student work that will reveal the critical differences in student representations and reasoning.
3. Use the ideas. A class discussion takes place that examines and debates student’s representations, reasons, and thinking. The teacher’s role is to pose questions and facilitate discussion that results in a consensus view(s).

A second adaptation to KF was to change the writing scaffolding prompts in the software to argumentation prompts. The argumentation prompts focused students on considering what counts as good and accurate data; the Argumentation 1 and 2 transformation points shown in Figure 1. Two probes of inquiry were used — one that asked students to consider the collection of data and one that engaged them in considering the analysis of data.

Following the completion of pulse data gathering investigations, students were presented with KF tasks to pursue students’ arguments concerning the collection of data. In the first activity students were directed to write and post a ‘note’ agreeing or disagreeing with each lettered statement below. The KF note window prompted students to supply a reason and to cite evidence to support an agree or a disagree position. The four data statements presented to students on the KF screen as notes are:

A. It matters where you take a pulse.
B. It matters how long you take a pulse (6, 10, 15 or 60 seconds).
C. It matters when you begin to take a pulse after exercising.
D. It matters who takes a pulse.

Upon completion of the write/post-a-note activity, we found that of the four statements, only the first two statements A and B produced a diversity of responses suitable for an assessment conversation on coordinating and conducting argumentation discourse activities. For this research report, our focus is only on Statement B - “It matters how long you take a pulse.” Examples of students posted notes (as written and submitted) for Statement B include:

- Note 1 It doesn’t really matter for how long you take the pulse, just as long as you multiply the pulse by the correct number of seconds.
- Note 2 I thought it does matter how long you take the pulse. The only true way to get the heart rate per minute is to take the pulse rate for a minute.
- Note 3 Yes, I think time does matter in taking a pulse. Such as being active. You stop being active and take your pulse. The longer you take, you heart beat will slow down to it normal rate.
The students were directed to open, read and respond to each of the notes. For each note, students were directed to use the KF scaffold tool: 1) to either agree or disagree with the statement; 2) to provide a reason for this decision, and 3) to cite evidence to support the reason.

The most frequently selected notes with accompanying arguments (reasons and evidence) can be found in Figures 2, 3, 4, and 5. The codes B-A-8 refer to statement B, Agree, and statement 8. The **boldface** words signify scaffold tool buttons we the researchers embedded into the KF environment. Students were instructed to use these tools as they wrote the agree/disagree statements. By design, as stated above, a diversity of student responses were selected for the purpose of studying what criteria students would use to assess and evaluate the position made by students in their statements.

Table 1 presents the outcome of the task. Three statements were judged by students to be the best. Namely, B-A-3 with eight votes; B-A-8 with six votes and B-A&D –1 with five votes. While the intention of the researchers was that students would make only one selection, many students followed the make-a-collection instructions and selected two or more best statements.

**Results/Points of View**

Consensus opinions about the best arguments did emerge and were

<table>
<thead>
<tr>
<th>B-A&amp;D-1</th>
<th>5 votes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statement B:</strong> It matters how long you take a pulse</td>
<td></td>
</tr>
<tr>
<td><em>Disagree</em></td>
<td></td>
</tr>
<tr>
<td><strong>Reason Why</strong></td>
<td>If you are working out and you want to take your pulse you have to do it right after you stop your activity.</td>
</tr>
<tr>
<td><em>Agree</em></td>
<td></td>
</tr>
<tr>
<td><strong>Reason Why</strong></td>
<td>If you are taking your resting heart rate it doesn’t matter how long you take it for because it will be consistent (sic).</td>
</tr>
</tbody>
</table>

**Reasons for Selecting B-A&D-1 as a Good and Accurate Statement:**

— You could really understand what they were talking about and saying.
— They explained themselves and I think they were right.
— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky.
— One of these explains agree really well with detail and another explains both pretty well.
— They are accurate and they prove the point well.

**Figure 2.** Reasons given for highly selected Best Statement B-A&D-1.

<table>
<thead>
<tr>
<th>B-A-3</th>
<th>8 votes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statement B:</strong> It matters how long you take a pulse</td>
<td></td>
</tr>
<tr>
<td><em>Agree</em></td>
<td></td>
</tr>
<tr>
<td><strong>Reason Why</strong></td>
<td>Because if you are recording the pulse after a certain amount of exercise because the blood pressure will start to go down after a certain amount of time, therefore making the number of beats per minute. It would be most accurate if you took the pulse for ten seconds multiplying it by six to find the number of beats per minute.</td>
</tr>
</tbody>
</table>

**Reasons for Selecting B-A-3 as a Good and Accurate Statement:**

— Give good support.
— It has a good point and backs up its point very well.
— It is clear and it explain the argument well (i.e. the blood pressure will start to go down after a certain amount of time).
— They have a good argument and because if you only take a pulse for two seconds and times it by 40 it won’t be accurate unless your lucky.
— They make their point. They brought new ideas and explained them. They also took old ideas but explained them better than the others.
— I agree with them and they prove they’re point well. They’re point is that if one takes one’s pulse for 3 secs. you would probably not find an accurate heart beat.
— They say that it does matter and then expalne why it matters.
— They are all making a different point and that makes more clear to understand the answers.

**Figure 3.** Reasons given for highly selected Best Statement B-A-3.
from the KF tasks. The data show that 12 students — the consensus group — choose at least one of the three top selected statements. Looking across the three statements and focusing on those students who opted for more than one choice, we see a further consensus among the students. That is, of the students making more than one choice many opted for one or the other of the top three selected statements. This is shown in brackets on the end of reason statements (e.g., [B-D-5]).

Following the lead of these 12 students, and with an eye toward establishing the context for the next assessment conversation on the issue of what counts as a good reason, the other selections of the consensus group were examined. Three of the consensus group students selected statement B-D-5, see Figure 5. This represents the only disagree statement. A careful read of the statement shows a contradiction by the student writer in that the narrative of the reason reflects agreement with statement B. This statement is also consistent with the other top selected statements in that it addresses the problem of the heart rate slowing down or calming down with time. Finally, it is the only statement among the consensus group that employed the evidence scaffolding tool. This is significant since a goal of the instruction is to model, scaffold and coach students into the practice of making arguments that site evidence.

We also found it interesting to look at the selection of statements from the student who choose seven and used the same reason for each of the selections - THEY HAVE A GOOD ARGUMENT AND BECAUSE IF YOU ONLY TAKE A PULSE FOR TWO SECONDS AND TIMES IT BY FORTY IT WON’T BE ACCURATE UNLESS YOUR LUCKY. The seven selections are found in Table 1. A review of the seven statements reveals a theme that is different from the other eight statements. That is, a concern for the magnification of measurement error when multiplying short durations of data collection (e.g., two or three seconds) to infer beats/minute. One can imagine how a subsequent assessment conversation can focus students’ attention on the role of measurement, the accuracy of measurement and the ways in which the spread of measurements can influence outcomes in an investigation. Such interrogation of data has been shown to be a critically important step toward students’ engaging in a deepening and broadening of explanatory models and mechanisms (Lehrer, Schauble & Lucas, 2008; Petrosino, et al., 2003; Lehrer & Romberg, 1996). What this program of research has established, and what we contend an exploration of reasons about reasons can reveal, is the confusion students typically make between the attributes and the properties of materials and organisms.

For the data collection task, the research data show that the majority of students are able to evaluate the quality of arguments. For 12 students a consensus of opinion emerged on what counts as accurate data for an exercising heart rate. We also see an alternative perspective on counting errors emerged with various suggestions on how to deal with counting errors. For example, recommendations to take the pulse for a full 60 seconds may work for resting heart rate but this strategy will not work for an exercising heart rate. With the passage of 60 seconds the pulse rate will slow down. Only some of the students see this situation and take it into account for the evaluation of the statements.

From Table 1 we can see that there were six statements that did not receive any votes.

These results suggest that students are able to evaluate the quality of arguments made by classmates and can do so employing a number of different reasons or perspectives. Disturbingly, the data also reveal that students are not providing evidence for the reasons selected. Students did not employ the KF evidence scaffold tool. This suggests there is confusion among students about the difference between reasons and evidence or the need to support reasons with evidence.
Table 1. Statement Selections and Reasons for Selection

**Statement B: It matters how long you take a pulse.**

<table>
<thead>
<tr>
<th>Statement B</th>
<th>Number of Students Selecting Statement</th>
<th>Reasons for Selecting Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A&amp;D-1</td>
<td>XXXXXXX 5</td>
<td>— You could really understand what they were talking about and saying;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They explained themselves and I think they were right (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— One of these explains agree really well with detail and another explains both pretty well (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They are accurate and they prove the point well (2).</td>
</tr>
<tr>
<td>B-A-1</td>
<td>X 1</td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7).</td>
</tr>
<tr>
<td>B-A-2</td>
<td>XX 2</td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— No reason given (2).</td>
</tr>
<tr>
<td>B-A-3</td>
<td>XXXXXXXXXX 8</td>
<td>— Give good support (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— It has a good point and backs up its point very well;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— It is clear and it explain the argument well (ie. the blood pressure will start to go down after a certain amount of time);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They make their point. They brought new ideas and explained them. They also took old ideas but explained them better than the others (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— I agree with them and they prove they’re point well. They’re point is that if one takes one’s pulse for 3 secs. you would probably not find an accurate heart beat (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They say that it does matter and then explain why it matters (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They are all making a different point and that makes more clear to understand the answers (3);</td>
</tr>
<tr>
<td>B-A-4</td>
<td>XX 2</td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— I agree with them and they prove they’re point well. They’re point is that if one takes one’s pulse for 3 secs. you would probably not find an accurate heart beat (2).</td>
</tr>
<tr>
<td>B-A-5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B-A-6</td>
<td>XX 2</td>
<td>— It says all you need to know and it uses a good example;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7)</td>
</tr>
<tr>
<td>B-A-7</td>
<td>XX 2</td>
<td>— Give good support (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They have a good argument and because if you only take a pulse for two seconds and times it by forty it won’t be accurate unless your lucky (7)</td>
</tr>
<tr>
<td>B-A-8</td>
<td>XXXXXXX 6</td>
<td>— They explained themselves and I think they were right (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— One of these explains agree really well with detail and another explains both pretty well (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— No reason given (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They say that it does matter and then explain why it matters (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They are all making a different point and that makes more clear to understand the answers (3).</td>
</tr>
<tr>
<td>B-D-1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B-D-2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B-D-3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B-D-4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B-D-5</td>
<td>XXX 3</td>
<td>— They are accurate and they prove the point well (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They make their point. They brought new ideas and explained them. They also took old ideas but explained them better than the others (2);</td>
</tr>
<tr>
<td></td>
<td></td>
<td>— They are all making a different point and that makes more clear to understand the answers (3).</td>
</tr>
<tr>
<td>B-D-6</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

*Number in ( ) indicates the number of times a student used the ‘reason statement’ to justify selecting a Agree/Disagree statement. For example, one student used the following statement seven times: THEY HAVE A GOOD ARGUMENT AND BECAUSE IF YOU ONLY TAKE A PULSE FOR TWO SECONDS AND TIMES IT BY FORTY IT WON’T BE ACCURATE UNLESS YOUR LUCKY.*
Educational/Scientific Importance
An understanding of how students choose and use evidence is a goal of this research study. Making thinking visible about choosing and using evidence is important to understand so that we can help to coordinate and focus teacher’s feedback on epistemic reasoning. Additionally, choosing and using evidence are core practices learners need to use when building and refining models.

Students’ using the analysis of the heart beat patterns and graphical representations to comment on issues about good and accurate data was successful to a point. We found that the development of the ability to distinguish reasons from evidence is an instructional situation that will need careful mediation from the teacher and from knowledgeable peers. It would seem that more attention needs to be given to the interrogation and modeling of the pulse rate data. Lehrer and Schauble (2006b) have reported that getting students to engage in resemble representation tasks is an entrée to modeling. Modeling, they maintain, is then sustained and extended by using resemblances to construct representational forms that afford quantification and investigation of relations among quantities (Lehrer & Schauble, 2006a; Lehrer et al., 2008).

There is also the possibility that the inability to distinguish reasons from evidence is related to confusions surrounding attributes (e.g., blood flow) and properties (e.g., pulse rate). The significance of this possibility is that domain-general rules for what counts as a reason, evidence, warrants and backings so not work effectively across different investigative contexts (Lehrer et al., 2008). Rather, such ‘what counts’ rules will need to be sensitive to both the conceptual and the epistemological domain under investigation. The development of learners’ abilities to talk about measurement, observations, evidence, patterns, and modeling will need frequent and insightful mediation from teachers and more knowledgeable peers.

This research study has raised questions and issues about the design of science learning environments that seek to promote scientific discourse practices. We see the following assertions and issues as relevant to the enterprise:

• The ability for students and teachers to reconfigure information allows post hoc discussions about evidence, methods of data collection, reasoning, and explanations. Computer supported discourse like KF allows the pace of argumentation to slow down and thus enable teachers and students to ‘take stock’ of the information and ideas emerging from scientific inquiries.

• Our interventions show that ideas and information emerging from small group and whole class activities can be captured by a computer server and used effectively to inform assessment decision making and the mediation of students epistemic reasoning.

• One of the main advantages of the computer supported environments like KF is that the visibility of student thinking is available for inspection and re-inspection by the teacher and students alike. Employing the KF platform in conjunction with other artifacts (e.g., notebooks, journals) makes possible studying the fate or outcome of a line of reasoning. We see very important implications for understanding the developmental corridors for learning epistemic practices of science. Knowledge Forum, in many ways, predates the structures and philosophies of “Web 2.0.” Web 2.0 is exemplified with wikis, blogs, and user-driven websites that value user-generated content. These environments include creating archives as collective memory, visualizing cognitive processes, and uniting groups that are geographically distributed (Andriessen et al., 2003; Zimmerman, 2005). Drawing on these theories about the nature of CSCL learning, researchers have reported on a number of structured online systems developed to facilitate individual and group science learning by emphasizing the formation of argumentative discourse through group conversations and writing projects (Bell, 2000; DeVries, Lund, & Baker, 2002; Scardamalia, 2002).

• Our study suggests that the ability to manipulate the metacognitive prompts (i.e., scaffold tool bar) is a powerful way to both engage and develop epistemic reasoning. When students become aware of their own learning processes, they gain much richer understandings of the content of their learning, and become better, more empowered learners (e.g. Baird, 1986). Metacognitive prompts can focus on students’ self-reflections of their interactions and roles, their self-perceptions of the learning tasks, and their learning strategies. Students can also be asked to explain and elaborate on any other strategies that they promoted their group and individual activities, learning, and task completion.

Acknowledgements
Thanks to our collaborating teachers Tim Hicks, Mike Kugler, and Tracy Klebe.

Bibliography


Driver, R., Newton, P., and Osborne, J., Establishing the norms


