Collaborative Inquiry into Students‘ Evidence-based Explanations: How Groups of Science Teachers Can Improve Teaching and Learning

Jessica Thompson, Melissa Braaten, and Mark Windschitl
University of Washington, Seattle Washington
Bethany Sjoberg, Evergreen High School, Seattle, Washington
Margaret Jones, Rainier Beach High School, Seattle, Washington
Kristi Martinez, Eastlake High School, Sammamish, Washington

Teachers in schools across the nation are gathering regularly with colleagues in Critical Friend Groups (CFGs) or as a part of Professional Learning Communities (PLCs) to examine student work and uncover how their young learners are thinking about science ideas. Such groups aim to 1) improve their practice by basing instructional decisions on evidence of student thinking and 2) improve outcomes for students. Achieving positive student learning outcomes, however, requires that teachers use principled ways of designing curriculum and instruction, selecting student work, analyzing these artifacts, and making evidence-based changes to instruction (NRC, 1996; Wood, 2007). Based on our research, we offer a model of collaborative inquiry for groups of science teachers who would like to systematically improve their practice through analyses of student work. We refer to this type collaborative inquiry as the Advancing High-leverage Practices by Examining Student Thinking, or the APEXST, model (Figure 1). The five components are elaborated upon as phases in the following sections. The APEXST model of collaborative inquiry has three distinct features: 1) a focus on a high-leverage practice (in this case pressing for evidence-based explanations), 2) a focus on longitudinal learning—both student and teacher learning over the course of a year, and 3) attention to students at all levels of the achievement spectrum.

Phase 1. *Collaboratively defining a vision of worthwhile learning*.

Perhaps the most important decision that teacher groups can make is to identify some aspect of student learning that is important enough to focus on for a full academic year (Curry, 2008; Windschitl, Thompson, & Braaten, 2009). We recommend teacher groups choose a high-leverage (Franke & Chan, 2006) scientific practice that can be developed over the course of a year in learners across different topics and even across different science courses. In our recent collaborative inquiry project with teachers we chose students' "construction of

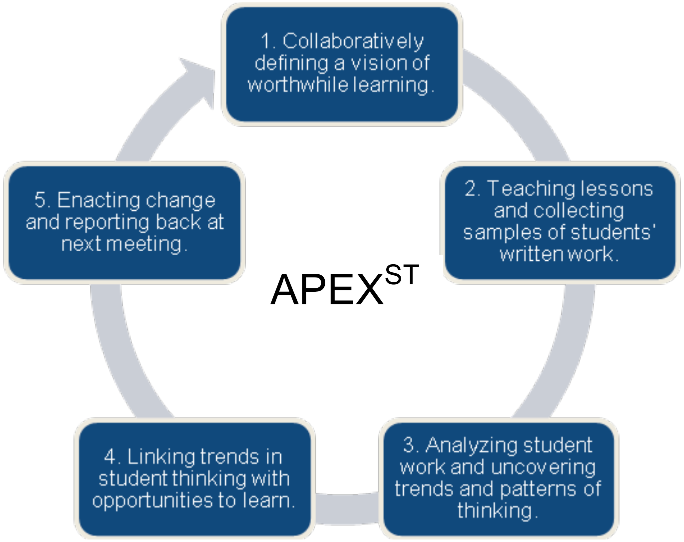

Figure 1. APEXST model of collaborative inquiry into practice

evidence-based explanations." This type of scientific thinking is critical to understanding the big conceptual ideas in science and it is a valued scientific practice (NRC, 2000; Author, 2008). We do *not* recommend that teacher groups choose student understanding of specific topics that change from meeting to meeting. This prevents a sustained opportunity by teacher groups to look at the development of an important scientific practice across time.

Developing lessons that press students for evidence-based explanations can be difficult intellectual work, and we have found that teachers benefit from discussing with peers what counts as a "big idea" and a rich scientific explanation. The challenge is that lessons from standard curriculum materials are often organized around topics or processes, not big ideas with rich underlying explanations. Big ideas have causal stories, composed of a web of events that help explain why observable phenomena occur. For example, one observable phenomenon could be diffusion of different materials across a membrane. The explanation for why this occurs has to do with equilibrium, concentrations of solutions, and permeability of membranes.

Once a big idea is selected, teachers can work together to identify a lesson or series of lessons that aim to explain observable phenomena. Typically lessons that ask students to synthesize data from an investigation or information from a body of existing evidence are the most fruitful for helping students wrestle with both the use of evidence and the development of an explanation. Examples might include students making sense of data from a pulley investigation by using ideas about forces, work, and energy to explain why a single person can lift a very heavy load using simple machines. Students could also synthesize what was learned from a series of investigations that examine air pressure in order to explain why molecules behave in predictable ways, in terms of kinetic molecular theory. Alternatively students could synthesize existing bodies of evidence such as DNA, fossil records, and morphology to explain examples of natural selection and evolution.

The next step for teachers is to detail a full explanation for the phenomena. This exercise may require looking up additional background information. A full explanation is a causal story that describes *why* a phenomenon occurs. There might be a chain of "why" explanations that complete the full causal story. To connect this explanation to evidence from an investigation, it is helpful to outline or diagram what is observable or measurable and then draw in features and processes that are not observable—thus creating a scientific model for the phenomenon under study (see Windschitl, Thompson, & Braaten, 2008 for more information on Model-Based Inquiry). Once the *why* explanation is outlined, the team of teachers can then write a rubric that details a *how* and *what* explanation (see row 2 in Appendix A. *Analysis of Student Understanding of Evidence-based Explanations*). Figure 2 provides an example of the *what-how-why* explanation framework that a group of teachers we worked with developed for a cellular respiration investigation in which students were asked to "Explain why you would see an increase in respiration after exercise." In the investigation students breathe into a Bromothymol Blue (BTB) solution before and after exercising as a direct indicator of carbon dioxide output and an indirect measure of glucose being converted to energy with carbon dioxide as a byproduct.

	Level 1	**Level 2**	**Level 3**
Depth of Explanation	o Student describes *what* happened. o Student describes, summarizes, or restates a pattern or trend in data without making a connection to any unobservable/ theoretical components.	o Student describes *how* or partial why something happened. o Student addresses unobservable/ theoretical components tangentially.	o Student explains *why* something happened. o Student can trace a full causal story for why a phenomenon occurred.
EXAMPLE explanation for cellular respiration investigation	The Bromothymol Blue changed color after exercise because the body exhaled more carbon dioxide as compared to when the body is stationary.	When exercising the body requires more oxygen. As oxygen intake increases so does the carbon dioxide output.	When exercising the body requires more oxygen which is taken from the lungs to muscle cells (via the circulatory system and diffusion). The cells use the oxygen to breakdown glucose into energy and carbon dioxide. Muscles use the energy to do work and the carbon dioxide diffuses into the blood and then the lungs and is exhaled. Cellular respiration happens at a faster rate when a person is exercising.

Figure 2. Row 2 of the Rubric for Examining 3 Dimensions of Evidence-based Explanations and a sample explanation co-developed by a group of teachers.

Teachers are now ready to co-develop written and spoken prompts to guide students toward "why" explanations.

To help students attend to the use of evidence to support such explanations, teachers can use *Row 1—Degree to which the student makes comparisons among pieces of evidence* and *Row 3—Degree to which evidence and explanations are integrated in written products* from the attached *Rubric for Examining 3 Dimensions of Evidence-based Explanations*. The rubric can be used both to design questions for students as well as to evaluate students' written work.

One of the groups of teachers we worked with wanted students to wrestle with multiple forms of evidence and to coordinate these with a scientifically rich explanation. They devised an investigation that followed the one described above, in which students collected three types of data before and after exercising: heart rate, number of breaths per minute, and amount of time in seconds it takes for BTB to change from blue to yellow. Following this investigation, students considered two additional forms of evidence in the form of written text. They read about how the brain controls carbon dioxide levels in the body as well as about what it means for the body to be physically fit in terms of lung capacity, blood vessels, and the heart. Students used these five "buckets of evidence" to answer these questions: "How and why does your body return to a resting rate after being at an elevated rate in terms of gas exchange and breathing rate? How might this be different for a person in good shape versus a person out of shape?"

Phase 2. *Teaching lessons and collecting samples of students' written work*

A key feature of the APEXST model is the attention to students of all achievement levels. We recommend selecting nine students to track throughout the school year—three high achieving, three average achieving and three underserved students. The chunk of work you analyze can be small, perhaps a 3-5 sentence response to an explanation-type question asked on a quiz, or it may be a bit longer if you choose to analyze an evidence-based claim students made following an investigation.

Many of our teachers have found it helpful to also collect video samples as a way to share images of rich classroom conversations with peers.[1] This is optional. As you might expect, recording and editing video samples is time consuming. If available, this work could be done by an additional support person—perhaps a district-level science coach, an administrator, or a research partner from a university. For video segments, try to capture *student* talk, and preferably student-to-student talk, as a way to provide insight into how students wrestle with evidence and/or scientific explanations.

Phase 3. *Analyzing student work and uncovering trends and patterns of thinking*

Most forms of assessment will tell you *if* students have learned, however principled examinations of student work can reveal *why* some students have learned and others have been less successful. This analysis phase of the APEXST model is usually done individually by teachers. The first step of analysis is to use the specialized rubric you've developed as a reference to detect what "partial understandings" students might have of the target understandings. This provides clues about different paths to understanding, rather than simply making judgments about whether students' responses were correct or incorrect. To conclude prematurely that a student is "wrong" or "didn't get it" distracts from the possibility that there is *some* learning taking place that you can build on.

Partial understandings by students may take several forms. Perhaps a student is familiar with a scientific idea, but uses it in the wrong context. Perhaps a student understands part but not all of an idea. Perhaps they recognize when a vocabulary term should be used but do not provide evidence that they understand what it means. Perhaps they can follow directions faithfully or go through the motions of a scientific practice (like interpreting data, or creating a model), but don't know when or how to apply this practice in a new situation.

Another aspect of analysis is to look carefully at the full range of students in your classroom—from high performing, middle, and struggling students. Attending to a full range of students' written work provides a foundation for teachers to address more systematically how a broad range of students are learning in the classroom. In our experience, teachers found important differences in how various groups of students interpreted classroom instructions, in how they approached particular tasks, and how they participated in scientific activities. One of the teachers working with us found that her struggling students were having difficulty writing scientific conclusions, not because they comprehended the task differently from their peers, but because they did not know how to respond to feedback the teacher had given them on this type of writing a few weeks earlier. The teacher realized she had to help her underserved students understand not only how to engage with scientific writing practices, but how to use feedback more productively.

---

[1] Examining samples of student work and video is considered to be part of the typical work teachers do. If a note to this affect is not part of current district policy or if the students' work or samples of video will be used to show to educators from other sites then we suggest distributing an opt-out consent form to be given to students and their parents.

In the first pass at student work, you can use post-it notes and highlighters to mark places where students show elements of understanding the target idea. Students who have similar responses to one another can be grouped together, and you can then look for one of two types of patterns in student thinking. One pattern is an unexpected *trend*—this is where many students responded to instruction in an unpredicted way. It signals perhaps that a learning situation did not present students with an opportunity to process ideas in the way that that you thought it did. Another pattern is a *relationship*. This is where groups of students, who share similar characteristics, perform similarly on a feature of the task or question. In one of our cases, the English Language Learners in a classroom were able to draw out sophisticated models of air pressure during small group work, but during whole class discussions, they were unable to verbalize their thinking to others. Also see Appendix B. *Ideas for Identifying Patterns in Student Work over Time* which can be used to analyze patterns in students' learning over time.

If you decide to include video as part of your analysis of student thinking, then here are some questions to ask yourself as you analyze the videotape and select a 5-7 minute segment:

- What did you hear your students talking about?
- What evidence is there for partial understandings?
- How does the video compare to patterns in student work?
- If students gave incomplete or puzzling responses, what questions might you ask to clarify or stretch their thinking?
- If students did not engage in talking about explanations or only gave "what" or "how" level explanations, what prompt or scaffolding could be used?

Phase 4. *Meeting with peers to link trends in student thinking with opportunities to learn*

In the APEXST model of collaborative inquiry, the actual team meeting should be characterized by three kinds of accountability. The first type of *accountability is to peers*, which means that you have all shown respect to one another by doing the "groundwork" before the meeting. For each presenter, this means having thoughtfully collected and analyzed student work, made copies for everyone, and is ready for an in-depth discussion. For other participants, being accountable to one another means reserving the full time in your professional schedule needed to engage in this process with others (typically at least 50 minutes for each presenter). Accountability to your peers includes being able to critique their ideas in fair and civil ways, not necessarily agreeing politely with every statement made.

The *second type of accountability is to the science itself*. As part of the collaborative meeting, everyone is asked to make public a shared understanding of what a full and complete explanation is for the phenomenon that student work was collected on. This "primes" all participants to see certain clues to understanding in the student work. It also forms the basis of a common language that can be used during the session. This is difficult work, but very necessary. We found that many times we were not confident of our understandings of some fundamental ideas in science, but we made these questions public because a level of trust in one another had been developed.

The *third type of accountability is to understanding the student work.* During the meeting, a segment of time is devoted to giving a generous review of the student artifacts, meaning that everyone looks for partial understandings, as if every paper, no matter how sparse, contains important clues to the workings of a student's mind. Participants need to resist glossing over a sample of work, only to dismiss it as "wrong." Participants should look within the work of each student, but also across the samples of student work. Are there similarities to how students responded to instruction? Are there stark and unexplainable differences? Does the work cause you to ask questions about the instructional moves made or the opportunities to learn?

It is helpful to avoid common pitfalls. It is remarkably easy when discussing student work to engage immediately in "repair talk"—that is, talk of how one would fix the instructional activity. This talk can be contagious; a skilled facilitator should steer participants instead toward understanding of student thinking.

There are two tools that are indispensable in maintaining accountability to science and to the student work. One of these is the *Rubric for Examining 3 Dimensions of Evidence-based Explanations* (Appendix A) and the specific explanation rubric teachers co-develop in Phase 1. The other is the *CFG Protocol* (Appendix C) used to structure the conversation with peers. Our research has demonstrated the importance of three specialized parts of the protocol: 1) the invitation to come to a group understanding of the best possible scientific explanation of the focal phenomenon, 2) the prompt to participants to seek out evidence of partial understandings, and 3) at the end of the collaborative session, the move for everyone to imagine how particular instructional choices made by the teacher may have influenced how the student responded to a question or task. This third part allows the group to hypothesize, *based on evidence*, how changes in instructional moves or changes in opportunities to learn could positively impact learning for some groups of students.

Materials needed for a Critical Friends Group (CFG) Meeting:

- Student work: The presenter should bring copies of 3-4 students' work (not all 9 students) that are particularly illustrative of the patterns he/she noted—circle the 1 or 2 assignment questions that were analyzed and include any notes/highlighting from the analysis. Bring one set of papers for each member of the team.
- Rubric: The presenter should bring copies of the rubric that the group co-constructed in the first meeting (based on Appendix A but specialized for the explanation at hand) and notes about all 9 students' thinking (Appendix D).
- Patterns & Questions: The presenter should bring written reflections on their analysis, summarizing 1) the patterns they saw in the data and 2) questions they would like their peers to focus on during the session.
- Protocol: The facilitator should bring copies of the CFG protocol for all members (Appendix C).

Phase 5. *Enacting change and reporting back at the next meeting*

For the APEXST model of collaborative inquiry, you will want to leave about 30-60 minutes at the end of each CFG meeting to cycle back to Phase 1. After the presentations, participants should take a few minutes to take notes about the experience—recording ideas about student thinking as well as "take aways," meaning ideas or practices that particularly resonated with a teacher. The strategies employed by the featured teacher(s) are labeled and recorded on a master list (i.e. Bethany: Using "buckets of evidence" to help students coordinate evidence for an explanation or Margaret: Using “back-pocket questions” to probe students’ understanding of the underlying scientific explanation.) This step makes the attempted strategies public and provides cohesion between meetings. Each teacher can choose one idea or practice to try out prior to the next meeting. Teachers can then work in pairs to discuss exactly which practices might be applied to an upcoming lesson and work through the activities listed in Phase 1.

Does this form of collaborative inquiry influence practice and student learning?

We have worked with groups of teachers who meet once and others who have negotiated with administrators to have three full professional development days per year for this work. The models we have run and their associated outcomes are listed below. Regardless of structure, most teachers made significant gains in how they conceptualized which scientific ideas are worth teaching, how they pose why-level questions in the classroom and how they identify and respond to students’ partial scientific understandings for all students in their classrooms. Thus, through collaborative inquiry, teachers were able to greatly improve students’ opportunities to engage in rich forms of scientific reasoning. Moreover, over time they grow into a community of ‘critical colleagues’ who were willing to hold each other accountable, to take intellectual risks, and to open up windows into each other’s classrooms.

APEXST Model 1- High stakes, high reward

- *Overall structure.* Teachers invest time and effort into three meetings per year. This option works well for teachers who have constraints about meeting frequently or for teachers who are devising a Professional Learning Community across multiple schools. We ran this model with a group of 11 first year science teachers.
- *Meeting structure.* For each CFG meeting teachers analyze 2-3 different assignments for each of the 9 focal students they selected. Larger groups of teachers are broken into groups of 3 or 4 so that each teacher has a full hour to present his/her analysis. At the end of the day, teachers re-convene and spend about one hour working on Phase 1.
- *Outcomes.* Based on our research, 7 out of 11 teachers who participated in this option dramatically shifted their teaching practice to press for “why level” explanations in classroom conversation and on assessments (based on classroom observational data) (Windschitl, Thompson, & Braaten, 2008 & 2009). Many of these teachers used a version of the explanation row from the attached rubric with their students and reported that their students also adopted a what-how-why framework for critiquing their own explanations. The other 4 teachers occasionally tried new practices in their teaching, but did not make substantial changes to their routines, at least not in their first year of the project.

APEXST Model 2- Regular influx of ideas and visions

- *Overall structure.* Teachers meet eight times per year. This option works well with groups of teachers who have regularly scheduled release time and/or who have the aid of a district-level science coach. The coach can pace the group by assisting teachers in designing lessons, videotaping/editing and evaluating student work for each meeting (see Nelson et al., 2008 for more information on coaching). We worked with two groups of teachers using this model. One was an evening video club with 10 teachers and district-level science coaches (see Sherin and van Es, 2009 for more information on video clubs) and the other group was a science department with 5 teachers plus one coach.

- *Meeting structure.* Each month one or two teacher(s) feature his or her analysis of student work (and possibly a video segment) as part of the CFG meeting. For the last 30 minutes of the meeting all teachers work in pairs to choose a strategy for an upcoming unit and commit to analyze related student work for the next CFG meeting. At the start of the next meeting teachers can report on patterns in student data according to a particular shift in practice. (In our experience typically about one quarter of the teachers are able to follow though on the enactment of a lesson and analysis of student work for the following CFG meeting.)
- *Outcomes.* Based on survey data, teachers involved in these groups reported most significant improvement in practice on: 1) posing more why level questions in the classroom and on assessments, 2) attending to students' partial understandings and 3) identifying a big idea worth explaining in science. They reported that the dimensions that best supported these shifts in thinking and practice were 1) having the opportunity to examine other teachers' student work, 2) watching video from other teachers' classrooms and 3) working with a supportive group.

NOTE: We tried an additional model in schools with small science departments (with 2 teachers) who met every two weeks during a common planning time. The benefit to working with smaller groups and meeting this frequently was that the teachers could readily recall the topics their colleague was teaching and the suggestions they made to improve instruction from the previous meeting. They were more invested in the outcome of evidence-based modifications in one another's classrooms. However, these small groups faced challenges. In one school a teacher stopped participating and the group was not large enough to continue meeting and in another school the group was not large enough to develop hypotheses about teaching and learning that could lead to shifts in practice.

Summary

The principled and collaborative analysis of one's practice for the purposes of improvement is the work of professionals. This inquiry itself is scientific, involving the generation of questions, producing and collecting evidence, and co-constructing theories about how and why students respond to instruction in particular ways. This iterative process should grow over time as teachers become more capable of defining what counts as valued types of scientific practices and what counts as competent student performances, become more proficient at collecting and analyzing student work, and become more capable of changing instruction in response to evidence. Through the APEXST model of collaborative inquiry it is entirely possible for committed groups of science educators to understand their students' thinking in new and deeper ways, and to eventually make evidence-based changes that bring science achievement within the grasp of all their students.

References

Curry, M. (2008). Critical Friends Groups: The possibilities and limitations embedded in teacher professional communities aimed at instructional improvement and school reform. *Teachers College Record*, *110*(4), 733–774.

Franke, M.L. & Chan A.G. (2006). *High leverage practices.* Retrieved May 12, 2009 from Inside Teaching, Goldman-Carnegie Quest Project Website: http://gallery.carnegiefoundation.org/insideteaching/quest/perspectives.html

Nelson, T.H., Slavit D, Perkins, M., Hathorn, T. (2008). A culture of collaborative inquiry: Learning to develop and support professional learning communities. *Teachers College Record, 110*(6), 1269-1303.

National Research Council (1996). *National Science Education Standards.* Washington, D.C.: National Academy Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning.* Washington, D.C.: National Academy Press.

Sherin M.G. & van Es, E.A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education 60*(1), 20-37.

Thompson, Windschitl, & Braaten (2008, April). *How pedagogical reasoning and practice develops across "Learning-To-Teach" contexts*. Paper presented at American Education Research Association, New York, NY.

Windschitl (2008). What is inquiry? A framework for thinking about authentic scientific practice in the classroom. In *Science as inquiry in the secondary setting*. (pp. 1-20). Eds. Luft, J., Bell, R., Gess-Newsome, J. NSTA press, Arlington, Virginia.

Windschitl, Thompson, & Braaten (2008). Beyond the Scientific Method: Model-Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Science Education, 92*(3), 1-27.

Windschitl, Thompson, & Braaten (2009, April). *Fostering ambitious pedagogy in novice teachers: The new role of tool-supported analyses of student work.* Paper presented at National Association for Research in Science Teaching.

Wood, D. (2007). Teachers' Learning Communities: Catalyst for Change or a New Infrastructure for the Status Quo? *Teachers College Record 109*(3), 699-739.

This material is based upon work supported by the National Educational Association Foundation, the Teachers for a New Era Project sponsored by the Carnegie Corporation, Annenberg Foundation and the Rockefeller Foundation, and the National Science Foundation under Grant No. DRL-0822016. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these organizations.

For more information about our work with teachers see: http://depts.washington.edu/mwdisc/
Or contact Dr. Jessica Thompson: jjthomps@u.washington.edu