DESIGNING HANDHELD SOFTWARE
TO SUPPORT CLASSROOM ASSESSMENT:
AN ANALYSIS OF CONDITIONS FOR TEACHER ADOPTION

William R. Penuel

Louise Yarnall

SRI International

This material is based upon work supported by the National Science Foundation under Grant No. REC-0126197. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ABSTRACT

Since 2002, Project WHIRL (Wireless Handhelds In Reflection on Learning) has investigated potential uses of handheld computers in K-12 science classrooms using a teacher-involved process of software development and field trials. The project is a three-year research and development grant from the National Science Foundation, and it is a partnership between SRI International and a medium-sized district in South Carolina, Beaufort County School District. In contrast to many recent handheld development projects aimed at developing curricular materials, Project WHIRL focused on the development of assessment materials. In Project WHIRL, teachers were asked to apply their own curricular materials, content understanding, and pedagogical content knowledge to the project. Teachers and SRI researchers, software developers, and assessment specialists worked together to design software and activities that could be used across a variety of topic areas and science and in multiple phases of instruction to improve classroom assessment. This design process revealed to the research team teachers' beliefs and assumptions about assessment as well as a wide range of practices they used to find out what their students know and can do, both informal and formal. In this paper, we focus on how teachers' initial teaching and assessment practices influenced the design of handheld software and the ways in which these designs have been used across a variety of teachers' classrooms. In addition, this paper provides some preliminary answers to two of the key research questions we outlined at the outset of our project:

- What kinds of software designs can be feasibly implemented in classrooms that support effective assessment practice?
- What are the conditions under which teachers can adopt handheld tools to support classroom assessment?

DESIGNING HANDHELD SOFTWARE TO SUPPORT CLASSROOM ASSESSMENT

The Significance of Classroom Assessment in Science

A broad consensus is emerging in American education around the importance of developing students' understandings of and abilities to conduct scientific inquiry (American Association for the Advancement of Science, 1993; National Research Council, 1996, 2000). To meet these new standards for teaching inquiry, teachers are adopting approaches to teaching science that actively involve students in extended projects and investigations (see, for example, Krajcik, 2001). A number of research efforts have been dedicated to developing and investigating curriculum materials and technologies to support inquiry-based science. To date, however, there have been just a few researchers who have investigated assessment materials to support inquiry science and almost none who have studied how new technologies might help expand teachers' range and improve the quality of their classroom assessments.

The effectiveness of inquiry-based science instruction in developing understanding turns in part on the use of effective assessment materials and tools. For example, White and Frederiksen (1998) report that students that participated in a reflective assessment activity as part of their physics inquiry curriculum turned in final project reports at a higher rate than students in control classes. By contrast, when students engage in hands-on investigations without tools for student self-reflection and comprehension monitoring, students often fail to see the "big ideas" behind the investigations (Barron et al., 1998; Petrosino, 1998). When teachers do assess student learning on their investigations, they often turn back to standard multiple-choice assessments that test recall of facts, which are a poor fit to the form and content of student projects (Means, Penuel, & Quellmalz, 2001; Young, Haertel, Ringstaff, & Means, 1998).

Improving classroom assessment practices has proven challenging for reformers, however. The school day may be organized such that teachers rarely have adequate time to plan assessment activities in a principled manner or learn new strategies for assessment from peers and experts (Black & Wiliam, 1998b; Darling-Hammond, Ancess, & Falk, 1995). Collecting more varied forms of data on student assessment is difficult for teachers. More data can produce "information overload" for teachers (National Research Council, 2001; White & Frederiksen, 1998). When new assessments are introduced, students may resist changes to the flow of classroom activities and the changes in what is expected of them (Black & Wiliam, 1998a). Different members of the school community—students, teachers, administrators, and parents—may hold views of learning that are inconsistent with innovative assessment practice; these views, moreover, are often difficult to change. Changing assessment practices, in short, challenges the core of teachers' identities and their strategies for solving the day-to-day dilemmas of teaching (Atkin & Black, 2003; Atkin, Sato, Coffey, Moorthy, & Thibeault, 2003).

Why Handheld Computers for Classroom Assessment?

Handheld computers offer some potential supports to facilitate broadening the range and frequency of teachers' assessment of inquiry science. Because they are computers, they make the gathering and aggregation of data for use by teachers easier to accomplish. Handheld computers are portable, a feature that has been exploited by researchers developing curriculum technologies to use in field investigations of their environment (Hsi, 2000; Novack & Gleason, 2001; Soloway et al., 1999; Tinker & Krajcik, 2001). Their portability means that assessment can be easily integrated into any phase of inquiry, anywhere in the classroom or in the field. They are low-cost, which means that for many classrooms it is feasible to make the technology accessible to all students and involve students actively in self-assessment.

A number of for-profit companies have developed handheld software programs to support traditional forms of assessment. Both Scantron and Kaplan, for example, have developed software for handhelds that allow students to complete multiple-choice and short-answer tests, either as part of their preparation for standardized tests or as part of formal classroom assessments. To create the tests, teachers can draw from the companies' vast item banks to construct their tests for students, which are downloaded to a student handheld computer. As students take their tests on the handhelds, the programs give students feedback about the correctness of their answers and the percentage of answers they got right. Teachers can view individual and aggregate results using a program on their desktop.

Some other companies have developed innovative observational assessments, which are intended for use by teachers. Sunburst Technology's Learner Profile to Go, for example, allows teachers to record observations of student behavior and to keep track of evidence of students' progress in meeting standards. Like Scantron and Kaplan, Sunburst has developed content resources for teachers to aid in constructing observational assessments. Wireless Generation's handheld assessment software simplifies the data capture and management process for elementary-level teachers who maintain running records of students' progress in reading. Their software allows teachers to capture evidence of students' developing reading fluency, ability to correct decoding errors, and comprehension. New products in mathematics assessment follow a similar model, providing the teacher with a handheld device to facilitate the collection and management of classroom assessment data. The handheld assessment technologies that have been developed so far have tremendous potential to make it easier for teachers to assess students more frequently and manage the data they collect.

There are additional ways that handheld technologies might be used to improve classroom assessment that not been widely explored, however. For example, none of these handheld technologies have sought to broaden the repertoire of teachers' classroom assessment strategies. Instead, the software described above all attempts to make it easier for teachers to engage in assessment practices that they already may do, whether it is administering multiple-choice tests or observing students' progress in reading. Broadening assessment strategies to include tasks would allow teachers to gather evidence about skills that are not easily tapped by multiple-choice tasks, such as students' ability to formulate a scientific question or to represent their understanding of a complex system visually. Existing assessment software for handhelds has also not yet explored ways to involve students more actively in self-assessment. Although in some cases, feedback is provided for students, that feedback does not require students to re-think their approach to a topic, to question a step they may have taken in a lab or project, or to check their own understanding of a topic by explaining something they've created to another student or to their teacher. Handheld computer software could be developed that supports each of these forms of involvement in reflection and self-assessment, especially because their low cost means that all students could participate actively in assessment activities.

Our Approach

In this paper, we describe the preliminary findings from a research and development effort that explored how handheld computers might support improved classroom assessment in science classrooms at the middle-grades. The project, Project WHIRL, is a three-year research and development grant from the National Science Foundation, and it is a partnership between SRI International and a medium-sized district in South Carolina, Beaufort County School District. In Project WHIRL, we adopted a process of co-design of educational technologies to support improved classroom assessment of inquiry-based science. The co-design approach has been used widely in recent initiatives to develop new curriculum and assessment materials in science (Atkin, 2001; Black & Harrison, 2001; Fishman, Best, Foster, & Marx, 2000; Shrader, Williams, Lachance-Whitcomb, Finn, & Gomez, 2001). In these initiatives, both the process and products of co-design have been analyzed as potential supports for improving instruction and assessment. In particular, the process of assessment co-design has helped orient teachers to pay closer attention to assessing the quality of student work products, encouraged teachers to try out new instructional strategies, and helped develop among teachers a greater understanding of what their students knew and could do (Black & Harrison, 2001; Shepard, 1997). The products of co-design have given students greater opportunities to participate in class (Black & Harrison, 2001), and in some cases have allowed students to demonstrate gains in learning (Shepard, 1997).

At the outset of the project, none of the teachers interviewed or selected to be part of the project were familiar with the research base that guided the development of researchers’ goals or committed to those particular goals. The project needed a way both to learn what teachers’ own goals for science education were in the district and to elicit teachers’ own goals for participating in our project as they came on board. At the same time, Project WHIRL needed a way for researchers to share their expertise and excitement about the promise of improved assessment for learning with teachers in a way that was respectful of teachers’ own experience and that built trust so that the software and activities might actually be tried out in the classroom.

We therefore decided to adopt a process of co-design that was sensitive to the different values and approaches of researchers and teachers directly involved in the project to develop the software and assessment activities in Project WHIRL. We decided to include teachers in the design process from start to finish, in an effort to increase the likelihood that the software developed would be usable to them and adaptable to real-world classroom contexts. The process was structured, moreover, to ensure that teachers, researchers, and developers each had multiple opportunities to express their needs, concerns, and hopes for what kinds of assessment activities the software might best support. The facilitators of design teams sought out teachers' ideas first, as a matter of principle, before sharing their own ideas or suggesting activities to the team. Researchers were asked to offer their expertise cautiously and to withdraw suggestions when they were met with too much resistance from teachers.

The Current Study

This study describes what we learned about the kinds of software designs that are adoptable in classrooms. We learned from two phases of our work: the process of co-design and the field trials. The first part of this paper will describe our co-design process and the methods we used to capture the process and resulting designs; the second part will describe the methods and preliminary results from field trials about the conditions under which teachers can successfully adopt the designs.

Learning from Co-Design: What Designs are Feasible to Enact

A key lesson from both our initial study of teachers' classroom practices and the co-design process was that for designs to be adoptable, they must "meet teachers in the middle." In other words, designs must not represent too great a departure from teachers' current practice even if researchers' aims are to support more reform-oriented, inquiry-based teaching practice. The co-design process gave researchers ample evidence that building a bridge too far would result in designs that could prove unusable to a wide variety of teachers. Our resulting software reflects the dual goals that we hoped to accomplish: to design tools that supported both broad-scale adoption and to broaden assessment practice. In this section, we provide a broad overview of the co-design process, the methods we used to study it, and a case study of a design team that illustrates the potential value of our approach.

The Co-Design Activities with Teachers

The co-design process borrowed elements from participatory design (Schuler & Tamioka, 1993), extreme programming (Wells, 1999), and co-design methods employed by SRI researchers on the Educational Software Components of Tomorrow (ESCOT) project (Roschelle & DiGiano, in press). It began with a design conference that brought together 6 selected teachers from the district, administrators, and SRI researchers and software developers. These teachers were selected from applications and interviews. Overall, they were mostly experienced teachers, with a median experience level of 21 years of teaching. All the teachers had advanced degrees. There were three elementary teachers, one middle school teacher, and two high school teachers. At the conference, three design teams were formed and a charter developed. In fall 2002, teams met by teleconference to develop scenarios of use (see Carroll, 1995) and to develop requirements for the software. Paper prototypes were then developed and tested in the classroom before programming began. Once requirements had been revised, the software developers implemented the designs of teams and a new round of testing began, beginning in winter 2003. This cycle then repeated 2-3 times (depending on the team) for the remainder of the school year.

During the course of the 2002-03 school year, Project WHIRL software teams developed 2 new pieces of software (*Boomerang* and *Data Doers*) and made enhancements to two existing pieces of software (*Sketchy* and *Quizzler*). The HOT-Q team developed *Boomerang* as a tool that was designed to help students generate questions from teacher prompts or “answers.” The Data Doers team developed *Data Doers* as a tool that could help students catch errors they might make while collecting or recording data from classroom laboratory experiments or fieldwork. The Image Makers team added, among other things, color and the ability to create backgrounds to *Sketchy*, software developed by researchers at the Center for Highly Interactive Computing in Education (HI-CE) at the University of Michigan. Finally, the HOT-Q team worked with John Covele of Pocket Mobility, Inc., to make modifications to their *Quizzler* software, which allows teachers to distribute, re-aggregate, and score multiple-choice quizzes to students on a Palm OS-based computer.

Research during the Co-Design Process

During this phase, other SRI researchers were involved in conducting observations in classrooms and interviews with design team participants. The focus of the classroom-based observational research was on identifying the local effects of the use of prototype versions of the software on the flow of classroom activities. We conducted a series of classroom observations in one classroom from each of the three design teams. These activities were conducted before Project WHIRL software was introduced, during the first introduction of prototype versions of the software, and after initial trials of the software. We sought to develop from these observations hypotheses about the effects Project WHIRL tools might have on student self-regulation in classrooms where the tools would be used. Interviews with design team participants were designed to collect empirical documentation over time about the evolution of teams over time. In Project WHIRL, an independent observer (not involved in the design process) interviewed each design team member in each of three design teams once every four to six weeks using a structured protocol. The protocol asked about contacts with other design team members, activities, decisions and progress, what was going well, concerns, attitudes toward the project and team, and perception of the participant's role in the project. Finally, project meeting notes were kept for each of the design teams that described discussions and key decisions made by each of the teams. In this report, the case study of one design team is reconstructed from a combination of project meeting notes, observations and reflective accounts of tests of early versions of the software conducted in two teachers' classrooms.

The HOT-Q Team: An Exemplar of Meeting Teachers in the Middle

The HOT-Q team was chartered at the summer design conference with two teachers, a software developer from SRI, and the first author of this paper. The team, initially called the "Flashcards" team, wanted to create what its name suggested: a program that presented students with electronic flashcards on handheld computers in order to help them develop better conceptual understanding. The SRI members of the team were initially reluctant to help this team out to meet its goals for two reasons: (1) we believed such review tools already existed for the Palm, such that expending effort to create a new one would not be worthwhile, and (2) we were concerned that the tool would not provide significant new roles for students in actively reflecting on what they know and can do.

From the outset, this team faced a difficult challenge in that the researchers' goals and the teachers' goals for the software were in conflict. The design process we devised and shared with teachers privileged teachers' construction of the problem space, but responsibility about how resources would be allocated to software development remained with SRI. This team ultimately struck a compromise that met teachers' current needs while also proving generative for supporting the assessment of more advanced inquiry skills. SRI supported teachers' adoption of both a "flashcard"-like review tool and the development of software that would enable teachers to help develop students' question-posing abilities, a key goal of science inquiry (National Research Council, 1996).

This creative compromise was reached only after the team wrestled with the kinds of classroom realities that the teachers faced. Both teachers on the team were elementary school teachers, and they had few breaks in their day for intensive lesson planning. Team meetings with SRI had to take place after school, when this husband-wife team was at home with their children. Both had more than 10 years' experience in teaching and were National-Board certified, but neither used computers to support instruction on a regular basis. Neither had ever used a handheld computer before they began their involvement with our project. They were therefore unfamiliar with the capabilities of handheld computers and initially had a hard time imagining what they might want to do with them. Neither teacher reported having adequate in-school technical support or the authority to ask for more help, so reliance on the school's network would need to be minimized. Although accepting this constraint meant we would rely on beaming rather than classroom networks for communication, we avoided problems caused by network down-time (see Tatar, Roschelle, Vahey, & Penuel, 2003, for a discussion of this issue).

We found clues to the teachers’ concerns with concept review from observations conducted in their classrooms. For example, Sarah, the fourth grade teacher, used frequent student questions to find out what her students know, but most of these questions required students simply to recall facts they had read in their textbook or heard Sarah review earlier. We observed Sarah ask students questions like, “What do scientists use to measure temperature?” and “When water is a gas, what do we call it?” Sarah asked these questions of the class as a whole in rapid succession, allowing for her to find out a little bit about what students know on lots of topics. Student questions, moreover, got little attention, especially when they were more open-ended and related to inquiry. For example, a student preparing their cellophane covered cup asked Sarah, “What if it rains?” Instead of using this as a starting point for a conversation about how to control conditions of an experiment, Sarah just said, “We’ll have to fix it, and adjust.” Sarah’s husband Shawn’s approach to classroom assessment was similar to Sarah’s. He even stated at the design conference that he saw little value in students’ questions, because he felt his students could not ask good questions.

Early on in the fall meetings, SRI researchers suggested that in addition to helping the teachers ask questions of their students, they might create a tool to support student question building. This suggestion was met with immediate ambivalence. Shawn liked the idea in principle, but his belief about his students’ capabilities led him to conclude that gathering student questions would not yield useful assessment information. Neither teacher saw an immediate need for focusing more attention on inquiry-oriented “wonderment” questions in science, in order to help students develop question-posing skills. Both teachers were primarily concerned with pressures to cover content that would be tested on the state science test, especially among lower-achieving students that they believed needed to focus on understanding definitions of key terms.

SRI researchers agreed to find an existing tool that supported the flashcard functionality desired by the teachers, and they quickly became proficient at using it. The *Quizzler* program we selected had the ability to rapidly distribute quizzes and then collect the data from students. Its functionality sparked the imaginations as to what was possible, and soon the team was co-developing a program with the company that developed *Quizzler* called *Gradebook,* which would allow teachers to integrate student scores and record the number of times students tried particular questions from quizzes into their own electronic grade books. Ultimately, the match with their needs also helped the two teachers become proficient in orchestrating handheld use with students in their class. In just a few short months, they went from being only occasional users of desktop computers with their students to using handhelds 2-3 times per week with students.

During this time, we also convinced the teachers to try out a tool that would allow them to collect and share student questions. Ultimately, it was not our own suggestions but students' responses to early prototypes of the software that convinced both teachers of the value of developing a tool to support student questioning. The teachers agreed to try, as a prototyping activity, to reverse (or "Boomerang") the teachers' notions of who asks the questions and provides the answers in a classroom. In this activity, teachers would beam to students the "answers" and then students would create questions to go with the answers in the Palm Memo Pad software. Then students would beam back their questions for whole class review. The teachers likened the process of collecting student questions in response to teacher prompts to the television game show "Jeopardy," a connection that reinforced a traditional notion of science learning as fact-based but that helped both teachers see how the software could help them meet their original goals and concerns.

After just one class with this activity, both teachers were impressed with the student questions and the classroom discussions that happened when the student questions were displayed. The experience revealed to them just how much skill their students already had in developing good questions; it helped to convince them that their students could improve the quality of their questions. Although the teachers' interest in the new tool was tenuous at first, it continued to grow as they tried out new versions of the software in their classrooms. By the time that new teachers were added to the project in spring 2003, Shawn had become a strong advocate for the software. In just four sessions, he told a design team teacher on another team, his students could be taught to pose really good questions.

The Project WHIRL Field Trial: Learning about Conditions of Successful Use

While designing the software with teachers, we paid somewhat less attention to preparing teachers explicitly to use the software in ways that reflected good formative assessment practice. We designed the software to support a wide spectrum of practices, including the teachers' own early intuitions about how to use it. We decided to worry less that initial awkward uses of the WHIRL software, but we also learned from the co-design process that if we expected to see effective assessment uses of the software we would need to seed new ideas about how the software might be used as we expanded the project to 11 new teachers who would participate in a field trial during the 2003-04 school year.

Several conditions have emerged this year as important to supporting assessment uses of our software. First, the designed professional development experiences provided teachers with classroom assessment frameworks and assessment use from which to draw. Second, teachers' own existing repertoire of teaching and assessment practices served as a base from which to build activities with the new handheld software. Third, teachers' ability to manage classroom activities was challenged by the introduction of handheld computers for every student; those teachers that brought skill in managing multiple groups simultaneously were particularly successful in integrating the software with instruction. Finally, easily accessible technical support—provided through the project—has been an important contributor to some teachers' successful adoption of the software. In this section, we provide an overview of the supports we provided to teachers in the field trial, the methods we used to study use of the tools during the field trial, and contrasting cases of how two teachers new to the project came to use both the flashcard program and *Boomerang* in their own classrooms.

Supports for Teachers during the Field Trial

All of the teachers in the field trial received a classroom set of handheld devices, a charging station, and access to a full-time technology-learning coordinator hired locally to provide technical and pedagogical support to teachers. In addition, all received training in the use of all of the project's software applications and professional development activities to introduce them to pedagogical strategies including fostering and using students' questions to understand what students know and can do. The designs for these activities were based in part on lessons learned from working with design team teachers, especially the need to "seed" ideas for using the software in ways that might promote inquiry.

There were multiple formats for preparing teachers to use the software, and design team teachers were involved in several professional development activities. First, new teachers encountered the software by using it as students would in the context of a curriculum unit several teachers in the project taught. Teachers used *Boomerang* in this activity to formulate testable questions they could investigate through classroom experiments using Wisconsin Fast Plants. In addition to these experiences, 7 of the 12 new teachers in the project have opted to enroll in a graduate course being offered as part of the project. The course is aimed at helping teachers understand the conceptual underpinnings of effective formative assessment. On the basis of the experience with the design team teachers working on *Boomerang*, one session was dedicated specifically to the role that eliciting student questions might play both as a source of assessment data and as an early step in a student investigation in science.

An important element of the field trial emphasized to teachers that their role was to test out the software and help identify novel and effective uses of the software. We asked them to take on the role of designers of assessment activities, and we provided them with the supports described above to do so. At the same time, teachers were not asked to use the software for a specified amount of time; rather they were asked to use the tools as frequently or infrequently as they saw fit. We hoped to use the natural variation in usage patterns in our research to understand the conditions for successful enactment of assessment use of the software.

Research Methods during the Field Trial

To understand the conditions that would support adoption of handheld tools, we organized our research around a general hypothesis that teachers would adopt WHIRL tools at different levels of frequency and sophistication depending on their pedagogical background, their attitudes about student learning and assessment, their existing classroom instructional and management practices, and their level of involvement in a range of professional development opportunities we would provide. To test our hypothesis, we designed a multi-method analytic study that featured:

- pre-test and post-test measures of teachers' and students' attitudes, beliefs, and practices for monitoring progress, learning, and understanding;
- monthly classroom observations (n=~80 classes, ~2-3 per class);
- monthly teacher logs and occasional emails regarding goals in using handheld tools, uses of handheld feedback, and lesson designs; and
- automated handheld log files tracking actual use of all tools across all classes.

We are still in the process of gathering and analyzing these data; results of our analysis of changes in teacher practices and student attitudes, beliefs, and self-regulation will be conducted in summer 2004.

Contrasting Cases of Use: Wendy and Pamela

The field trial teachers were selected through an application process conducted in early spring 2003. The 9 elementary teachers and three middle school teachers then joined the project in late spring and committed to participation in the field trial with the written support of their principals. As a group, they were less experienced than the design team, with a median of 2.75 years in teaching, compared with over 20 years teaching among the design team. About 15 percent of teachers held advanced degrees. They also came from schools with a varied mix of student racial, ethnic, socioeconomic, and achievement characteristics.

Our new field trial teachers were similar to the design team teachers in one respect: their initial assessment practices. Our pre-surveys of teachers showed that all teachers (both co-design and field trial) initially engaged in a wide range of assessment practices, but tended to focus on assessing students for accountability purposes rather than assessing students to understand their thinking. Teachers also favored informal practices of assessment—such as directly questioning individual students—over more formal practices of quizzing and testing. These would prove an important resource or bridge for teachers' understanding of the project, as many of them came to embrace the tools for their ability to support what they termed "informal assessment."

Wendy and Pamela, two different field trial teachers, are both experienced teachers but have used the software in different ways and experienced quite different levels of success. Wendy is an eighth grade science teacher in a middle school in the wealthier southern part of the school district. She has struggled whenever she uses the software in her classroom, and her students have only infrequent opportunities to use the software. By contrast, Pamela, a fourth grade teacher in the elementary school where Shawn and Sarah teach, has made extensive use of both *Quizzler* and *Boomerang.* She finds it easy to integrate the tools throughout her curriculum and to orchestrate activities with her students.

The two teachers have taken advantage of the professional development opportunities offered through Project WHIRL at different levels. Wendy attended the three required workshops, and has called on the local learning technology coordinator for technical support. By contrast, Pamela participated in the workshops and took a more active role in them, presenting to her colleagues in February her methods for orchestrating handheld use in the classroom. She also completed the graduate course, where, by her own account, she learned ideas about how to incorporate student questions into her curriculum at the beginning of units. The result was that Pamela appeared to her colleagues and members of the research team as more invested in the project and as a much more central figure within it.

Both teachers' existing teaching and assessment practices shaped strongly how they used the tools. Wendy considers herself a constructivist teacher who wants to engage students in creative activities. She attempted to use *Boomerang* to engage students in a process of creating test questions and categorizing the kinds of questions they authored according to Bloom's taxonomy. However, whenever she attempted to use the tools according to their intended purpose, her difficulties with classroom management (described below) hampered their effective use.

Pamela's process was different. One practice that was familiar to Pamela at the outset of the project is something that is commonly called a K-W-L (Know-Want-Learn) activity, an instructional technique developed by Ogle (1986) that is designed to elicit, among other things, what students want to know about a topic. Pamela quickly began to use *Boomerang* before every unit to elicit what students know about a topic and what questions they have about it. She then uses this information to guide the course of the unit and re-visits student questions throughout the unit to see which ones have been answered. She also uses *Quizzler* nearly every day, because, like the pencil-and-paper quizzes she gives, the tool helps her to gauge what concepts students are struggling with and adjust the pace of her instruction accordingly.

The two teachers differ strongly in the degree to which they have been successful orchestrating use in the classroom. Observers to Wendy's classroom report that the time required for set-up is much longer than in other classrooms. They also report that students often appear confused, and when there are technical difficulties Wendy is quick to abandon her plans to use the software. Interestingly, students appear similarly confused in Wendy's other classrooms where she is not using handheld computers. Small group activities rarely run smoothly, and instructions are rarely clear to students in labs. By contrast, visitors to Pamela's classroom report that her classes are a model of how to organize instruction with handheld computers. She has developed a distribution and collection system for both devices and information that runs smoothly when she is using the handhelds. Her students appear to understand what is expected of them, and when technical difficulties arise, only rarely do they disrupt her plans so that she has to give up using the handhelds. Other, non-handheld based activities run just as smoothly in Pamela's classroom, providing some evidence of a more general ability to manage classroom activities so that the purpose is transparent to students.

Both of the teachers have drawn upon the readily available pedagogical and technical support, but the focus of their requests for help has been different. Wendy has sought primarily technical support to help her address what to her seems like a bewildering array of technical problems associated with her handheld computers. Nearly all of the problems have been addressed within 1-2 days by the local learning technology coordinator, but to Wendy, they often seem to be overwhelming. By contrast, Pamela has drawn on both the technical and pedagogical expertise of the local coordinator and her colleagues in the school, Sarah and Shawn. The local coordinator has been especially helpful in providing her with assistance as she experiments with implementing science kits provided by the state, which she says has helped her see herself as a more capable science teacher.

Conclusion

When we began our project, we knew that we needed to incorporate understandings of actual classroom practices and teacher pedagogical beliefs into the design of handheld tools. Through our work, we have developed a clearer understanding of the specific sorts of classroom practices and teacher beliefs that influence adoption of handheld technologies.

Our first question that we answered from the co-design phase of our work focused on defining the kinds of software designs that could be feasibly implemented in classrooms to support effective assessment practice. We have found that teachers favor software designs that are easy to learn and that map onto their existing repertoire of instructional and informal assessment practices in a highly fine-grained way. For example, we found that teachers identified specific concerns about student learning and sought to use software to address those concerns. They wanted to know better how many students understood from their reading and from lectures basic science vocabulary and ideas more frequently so they could adjust their instruction. They wanted to engage students personally in science and found question generation and animation provided ways to address those needs. They wanted to ensure students were collecting data thoughtfully and critically, and they wanted tools to assist.

We also learned about the need to temper researcher expectations of how much new technology can import reform practice into a classroom. If researchers seek to use software to extend or reform teacher practices, we learned that one possible route focuses on selecting focused additions to teachers' inquiry-oriented practices. We found that increasing students' participation in generating questions and getting students to think critically about data collection were two successful routes. The animation tool appealed to teachers too as a way to motivate students to study and represent their science knowledge in a new medium; it mapped well onto existing practices of making PowerPoint presentations and creating diagrams, charts, and posters on paper.

The second question that informed our field trial phase focused on understanding the conditions that could support teachers adopting handheld tools to support classroom assessment. Here were found that to ensure frequency of use, designers can benefit from focusing carefully on existing teacher practices and ensuring that technical support is easily accessible to teachers. Secondly, we are beginning to understand more about the types of teacher skills that can lead to higher quality adoption. We have found that teachers with strong classroom management skills and openness to reflecting on instructional practice with fellow teachers tend to use the tools and extend the use of tools in ways consistent with educational reform in science.

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Atkin, J. M. (2001, April). *How science teachers improve their formative assessment practices.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Atkin, J. M., & Black, P. (2003). *Inside science education reform: A history of curricular and policy change*. New York: Teachers College Press.

Atkin, J. M., Sato, M. D., Coffey, J. E., Moorthy, S., & Thibeault, M. D. (2003). *The local and the practical: Exploring teachers' assessment practices.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., et al. (1998). Doing with understanding: Lessons from research on problem and project-based learning. *Journal of Learning Sciences*, *7*(3-4), 271-312.

Black, P., & Harrison, C. (2001). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review*, *82*(301), 55-61.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*.

Carroll, J. M. (1995). *Scenario-based design.* New York: Wiley.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work.* New York: Teachers College Press.

Fishman, B. J., Best, S., Foster, J., & Marx, R. W. (2000). *Fostering teacher learning in systemic reform: A design proposal for developing professional development*. Paper presented at the National Association for Research in Science Teaching Meeting, New Orleans, LA.

Hsi, S. (2000, April). *Using handheld technologies to connect Web-based learning to outdoor investigations.* Paper presented at the National Association for Research in Science Teaching Annual Meeting.

Means, B., Penuel, W. R., & Quellmalz, E. (2001). Developing assessments for tomorrow's classrooms. In W. Heinecke & L. Blasi (Eds.), *Research Methods for Educational Technology.* (Vol. Volume One: Methods of Evaluating Educational Technology.). Greenwich, CT: Information Age Press.

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

National Research Council. (2000). *Inquiry and the National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. (2001). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academy Press.

Novack, A. M., & Gleason, C. I. (2001). Incorporating portable technology to enhance an inquiry, project-based middle school science classroom. In R. F. Tinker & J. Krajcik (Eds.), *Portable technologies: Science learning in context* (pp. 29-62). New York: Kluwer Press.

Petrosino, A. J. (1998). *The use of reflection and revision in hands-on experimental activities by at risk children.* Vanderbilt University, Nashville, TN.

Roschelle, J., & DiGiano, C. (in press). ESCOT: Producing educational software from reusable components by coordinating the reciprocal influence of research and development and classroom realities. *Interactive Learning Environments*.

Schuler, D., & Tamioka, A. (Eds.). (1993). *Participatory design: Principles and practices.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shepard, L. (1997). *Insights gained from a classroom-based assessment project*. CSE Technical Report 451. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Shrader, G., Williams, K., Lachance-Whitcomb, J., Finn, L.-E., & Gomez, L. (2001). *Participatory design of science curricula: The case for research for practice*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Soloway, E., Grant, W., Tinker, R., Roschelle, J., Mills, M., Resnick, M., et al. (1999). Science in the palm of their hands. *Communications of the ACM, 42,* 21-26.

Tatar, D., Roschelle, J., Vahey, P., & Penuel, W. R. (2003). Handhelds go to school. *IEEE Computer, 36*(9), 58-65.

Tinker, R. F., & Krajcik, J. (Eds.). (2001). *Portable technologies: Science learning in context*. New York: Kluwer Academic Press.

Wells, D. (1999). *Extreme programming: A gentle introduction.* Retrieved September 27, 2002, from http:www.extremeprogramming.org.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling and meta-cognition: Making science accessible to all students. *Cognition and Instruction, 16,* 3-118.

Young, V., Haertel, G. D., Ringstaff, C., & Means, B. (1998). *Evaluating Global Lab curriculum: Impacts and issues of implementing a project-based science curriculum*. Menlo Park, CA: SRI International.