J·T·L·A

The Journal of Technology, Learning, and Assessment

Volume 3, Number 5 · February 2005

Designing Handheld Software to Support Classroom Assessment: An Analysis of Conditions for Teacher Adoption

William R. Penuel & Louise Yarnall

www.jtla.org

A publication of the Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College

JTLA

Volume 3, Number 5

Designing Handheld Software to Support Classroom Assessment: Analysis of Conditions for Teacher Adoption

William R. Penuel and Louise Yarnall

Editor: Michael Russell
russelmh@bc.edu
Technology and Assessment Study Collaborative
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Design and Layout: Thomas Hoffmann
Copy Editor: Rachel Kay

JTLA is a free on-line journal, published by the Technology and Assessment Study Collaborative, Caroline A. & Peter S. Lynch School of Education, Boston College.

Preferred citation:

Penuel, W. R. and Yarnall, L. (2005). Designing handheld software to support classroom assessment: An analysis of conditions for teacher adoption. *Journal of Technology, Learning, and Assessment, 3*(5). Available from http://www.jtla.org

Notes:

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 authors and do not necessarily reflect the views of the National Science Foundation.

Special thanks to Nikki Shechtman at SRI for her assistance with data collection in the field trial and for preparation of figures for the paper.

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?

J·T·L·A

Designing Handheld Software to Support Classroom Assessment: An Analysis of Conditions for Teacher Adoption

William R. Penuel
Louise Yarnall
SRI International

Introduction

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 mClass 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 have 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.

The Current Study

In this paper, we describe the preliminary findings from the second and third phases of a three-phase 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 Beaufort County School District, a medium-sized district in South Carolina,. Each year of the project corresponded to a distinct phase of development. In the first year, we employed rapid ethnographic techniques to understand teachers and their work contexts (see Penuel, Tatar, & Roschelle, 2004, for findings of this phase of research). In the second year, we engaged a small group of teachers in the process of co-design of new handheld software and in testing this software in their classrooms. In the third year, we recruited additional teachers from the district to participate in a field test of the software and provided them with professional development to enable them to use the software effectively in their classrooms.

This study describes what we learned about the kinds of software designs that are adoptable in classrooms. We explore through case studies of selected teachers in our project what we learned from the latter two phases of our work: the process of co-design and the field trials. The first part of this paper will describe one teacher's experience with our co-design process and how his participation influenced the software designs; the second part will describe how two teachers used the software and engaged in professional development as part of the field trial of the software. The teachers depicted here are only partially representative of the teachers in our project. We selected these teachers and drew upon survey, interview, and observation data from these teachers to help illustrate the key benefits of challenges of incorporating handheld technologies to support classroom assessment.

The case study evidence presented here suggests that excitement about the potential of handheld computers to enable more frequent and effective formative assessment must be tempered with the recognition that teachers must have multiple opportunities to learn from peers and experts about how to adapt tools and manage their use in their classrooms. On the one hand, the cases lend evidence to the idea that tools that are usable in real classroom contexts and meet teachers' own perceived needs of teachers can be implemented successfully. At the same time, the cases also illustrate that handhelds alone cannot overcome teachers' challenges to engaging in more frequent formative assessment practice; they need access to others' ideas about ways the software can be integrated into their curriculum and strategies for orchestrating classroom use of handheld software.

Methodology

Participants

The participants in both the design process and in the field trial were all teachers in the Beaufort County School District, which was a partner in our project. Both groups of teachers were selected through a competitive application process. SRI staff made presentations throughout the district to principals and teachers in the 2–3 months leading up to the application due date. Teachers who were interested completed a 2-page application made available on the Project WHIRL Web site. The application asked teachers about their backgrounds, experience with technology, and approaches to instruction. No specific requirements for technology proficiency were designated, and the goal was to produce a group of teachers that was diverse with respect to prior experience with technology and approaches to teaching. For the design team application process, 10 teachers applied, and all were accepted by a joint team of SRI researchers and Beaufort district staff who reviewed the applications. Only 7 were able to attend the week-long design conference to kick off the project. For the field trial, 14 new teachers applied and all who met eligibility requirements were accepted. Two teachers fell outside the grade range targeted by the project and were not accepted.

For their participation, each teacher received a classroom set (between 25 and 30) of handheld computers and a charging station for use in their classroom. In addition, each year teachers received a stipend of $525 to cover their time participating in either design activity or professional development. Field trial teachers were given the option of taking a graduate-level course at the University of South Carolina for free instead of accepting their stipend. These teachers were required to participate in additional professional development as described below.

The teachers did serve a broad range of grade levels (4–9) by design: we sought to reach teachers who served students in transition to and from middle school, since National Science Education Standards denote important shifts in expected student understandings at each of these milestones (see National Research Council, 1996). A total of 7 teachers participated in the design process and 18 were part of the field trial. Table 1 shows the grade levels taught by teachers involved in each phase of our project. Note that the field trial teachers include all but one of the teachers who participated in the design phase of the project.

Table1: Grade Levels Taught by Teachers in Project WHIRL

Grade	Design Team Teachers	Field Trial Teachers
4	1	4
5		4
4 and 5	2	4
6		
7		2
8	2	3
9	2	1
Totals	7	18

The teachers varied with respect to their experience and preparation for science teaching. The median number of years teaching for the group as a whole was 7 years; the median of the design group was higher than the overall average, 21 years. The design team teachers who responded to our survey at the beginning of the field trial teachers all held a master's degree, but only 2 of the 11 teachers added during the field trial had more than a bachelor's degree. In both groups, nearly all the teachers held a K–8 multi-subject credential. No teachers had emergency or temporary teaching credentials. On survey scales designed to measure teachers' confidence in teaching the subjects they were required to teach, design team teachers did not differ significantly from field trial teachers. Both expressed generally high levels of confidence, except with respect to physics and teaching students who are English Language Learners (ELL).

Few of the teachers had extensive experience with integrating technology into their instruction or with using handheld technology. Initially, just 2 of the design team teachers used technology with students at least once per week. By 2003, 3 of 6 who responded to the survey indicated they used technology that regularly. By contrast, just one-third of the new field trial teachers used technology at least once per week. Although just two of the design team teachers had used handhelds before joining the project, two-thirds of the new teachers added during the field trial had used handhelds before. All teachers' knowledge and familiarity with handheld functions, however was limited. Most reported they could turn the handheld on and open up an application, but only 4 said they knew how to use Graffiti™ to write directly on the handheld screen, and 9 knew how to HotSync a handheld device to a computer to exchange and save data.

Data Sources for the Current Study

Project WHIRL collected a wealth of data from each teacher and her or his classroom as part of our research and development efforts. As part of the design phase of the project, we engaged teachers as participants in the research effort with us. Teachers and researchers together developed documents about the design requirements that are incorporated and described here. Each month, a researcher at SRI interviewed each member of the design team to learn about their ongoing experiences and impressions of the design process. From design team teachers' classrooms, we collected both videotapes of students using the software and conducted ethnographic observations of those classrooms when the software was being tested. Teachers added their own accounts of the software testing, which were recorded in minutes of the design teams. As part of the field trial, we gathered survey data from teachers in spring 2003 and spring 2004 to learn about their backgrounds, assessment goals and practices, and instructional practices. In addition, we conducted 64 structured classroom observations of Project WHIRL teachers using the software. Finally, we conducted interviews with half the teachers in the project in spring 2004, to learn more about what they perceived to be the benefits and challenges of participating in Project WHIRL and how they perceived using the project-supported software for assessment.

This research reports on interview and observation data collected from four of those teachers, as well as design team documents from the team to which two of these teachers belonged. We selected these teachers as the focus of this research purposefully to illustrate both the promises and challenges of using handheld computers to support classroom assessment. The two teachers selected from the design phase are both elementary level teachers who had limited experience with technology at the outset of the project but became avid users of handheld software for assessment through their participation in the project. The two teachers selected from the field trial phase were new to the project in that phase. Both were experienced teachers, one a middle school teacher and the other an elementary school teacher. One, however, was far more successful in integrating the tools into their assessment practices than the other teacher. Our analysis of the data from interviews and observations explore some reasons why these teachers differed in their success in adopting the software.

Design, Development, and Initial Testing of the Project WHIRL Tools

A key feature of our project was the process of engaging teachers in co-design of technology-supported assessment activities. The co-design approach we employed in this project has been used widely in recent initiatives to develop new curriculum and assessment materials in science (Atkin, 2001; Black & Harrison, 2001; B.J. 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).

Phases of the Project WHIRL Co-Design Process

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. 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. SRI facilitators of the design process 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 co-design process began with a conference that brought together the 7 teachers selected from the district, school district administrators, as well as SRI researchers and software developers. 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 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. A more detailed description of the design process and how it was experienced by different types of participants (designers, researchers, or teachers) will be available in Penuel, Roschelle, and Shechtman (in preparation).

Description of Software Applications Developed by Design Teams

During the course of the 2002–03 school year, Project WHIRL software teams developed 3 new software applications (*Boomerang*, *Data Doers,* and *Gradebook*) and made enhancements to two existing software applications (*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 (now owned by GoKnow, Inc.). Finally, the HOT-Q team worked with John Covele of Pocket Mobility, Inc., to make modifications to their *Quizzler* software and helped build a companion program called *Gradebook*, which together help teachers to distribute, re-aggregate, and score multiple-choice quizzes which students performed on a Palm OS-based computer. Below, we provide descriptions of each of the applications and how they are intended to support assessment.

Data Doers

Although students may believe that teachers do have "eyes in the back of their heads," teachers cannot be everywhere in the classroom at once. Consequently, from time-to-time students engaged in hands-on work may become distracted or confused about the task at hand and its relationship to the larger point of the experience. Especially during laboratory activities, students have basic questions about the activity: "What were those numbers?" "What was I supposed to do with them?" "What are we doing anyway?" Their confusion may last only a few minutes, and yet in the fast-paced world of the classroom, that few minutes can put them at a disadvantage.

Data Doers allows teachers to create handheld-based worksheets for labs or demonstrations to help students with data collection activities. *Data Doers* reminds students to think about what they are doing in two direct ways:

(1) Based on teacher-set upper and lower bounds for measurements, it gives students feedback about when a result that they enter is not plausible and needs to be reconsidered and possibly re-measured.

(2) It allows students to beam their data to each other, thus enabling them to make comparisons and contrasts more quickly.

It also provides more occasions for student thought in three *indirect* ways:

(1) Teachers can collect student values and respond in a more timely fashion than with paper-based systems.

(2) Students do not need to copy data tables during class discussion.

(3) Teachers can use the *Data Doers* spreadsheet to stimulate classroom discussion about a lab and its goals before the lab actually begins: a term the teachers on the team call "pre-flecting".

The flow of a typical *Data Doers* session is as follows:

Creating or Choosing a Lab. *Data Doers* allows classrooms to organize sessions by "labs" that are focused on a single hands-on activity (Figure 1). If the teacher needs to create separate versions for each class period, these can all be stored on the same device.

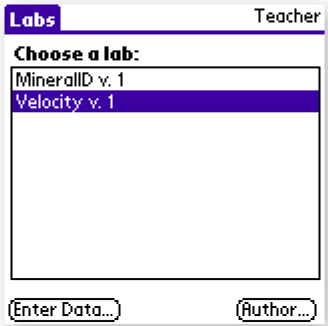

Figure 1. *Data Doers'* Screen for Choosing a Lab

Authoring. The first step in creating a Data Doers session is to create variables of interest and to define expected units and the range of expected values for each variable (Figure 2). If the author is not able to anticipate expected ranges, maximum and minimum values for variables can be left blank. In addition, the author identifies the values they wish student groups to compare when students beam results to one another or to a central "collector" device.

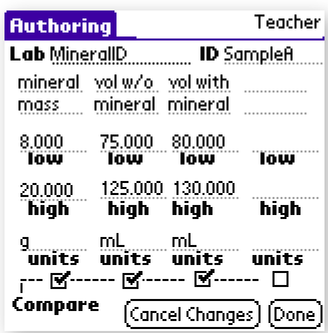

Figure 2. *Data Doers'* Authoring Screen

Distributing the Lab. Blank teacher labs can be sent from any device to any device using built in infrared beaming technology. Even if data have already been entered, the teacher or any student can beam a teacher lab with the authored ranges included in a blank version (Figure 3). In addition, the teacher has the option to beam the lab only, data only, the data set for the entire class, or to import data.

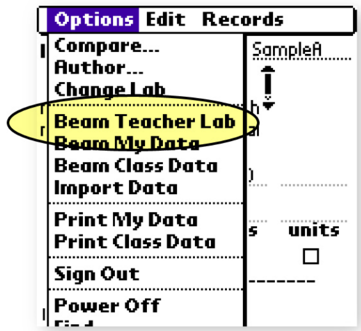

Figure 3. *Data Doers'* Beaming Option

Students Enter and Check Data. Students working individually or in groups record their measurements directly into *Data Doers*. Once they have entered the measurements, they can check their data by clicking on the checkbox below each variable. The *Data Doers* program either gives them feedback that there is a problem with units, the data are out of range, or puts a check in the box if their values are within authored ranges and have the correct units (Figure 4).

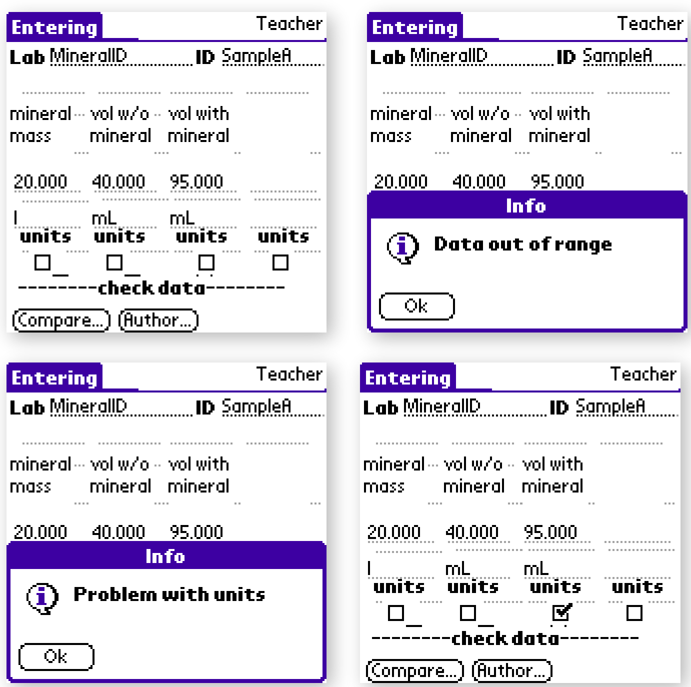

Figure 4. *Data Doers'* Entry Screen and Feedback Messages

Students Share Data by Beaming. Once students have entered and checked their data, they can beam their data to peers or to a collector device. If all data are gathered to a collector device, the class data can be re-beamed to all students so they can see all groups' data (Figure 5).

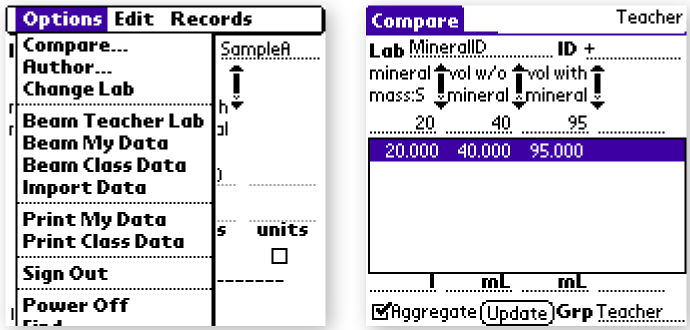

Figure 5. Screens Showing *Data Doers'* Beaming Data Function and Shared Data Displays

Teacher Projects the Data Table for Group Discussion. Using a document camera made available to all Project WHIRL classrooms, a teacher may display the compare screen and facilitate a group discussion of the data.

Boomerang

When students ask questions in their own words, they reveal gaps in understanding that may not be elicited by the teachers' use of standard terminology and phrases. Through question asking, not only do students fill in gaps in their knowledge base, but they also open up possibilities for wonderment. Standard classroom practice may permit only a small number of student questions. Often students with questions that might prompt important discussion or reveal a misconception are discouraged because other people's questions are so different from theirs.

Boomerang is a tool to support students asking questions. It allows all students to submit questions privately which can that can then be posted and discussed by the group as a whole. Students can record a question at any point in a lesson, unit, or hands-on activity. Teachers use Boomerang at the beginning of activities, to find out what students know about a topic or to help students formulate questions to guide inquiry. They use it in the middle to identify emerging problems in student understanding. And teachers use it at the end of an activity or unit, to help students review concepts or generate questions for a test.

The flow of a typical *Boomerang* session is as follows:

Teacher Creates Lists of Students within Classrooms. *Boomerang* allows teachers to create class names and lists of students embedded within classrooms. This function allows teachers to use the handheld device with multiple classes and students. It also helps teachers associate questions with particular students or groups of students (Figure 6).

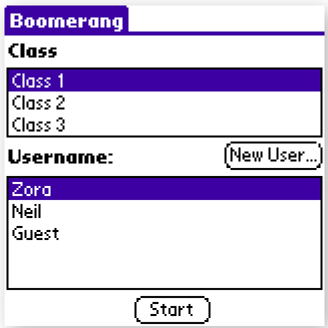

Figure 6. *Boomerang's* Class List Screen

Teacher Decides on Prompts for Student Questions. The classroom teacher provides prompts for student questions. These prompts are organized into chapters and subtopics (Figure 7). Prompts could be topics to be studied, concepts, or "answers" for which students might generate questions.

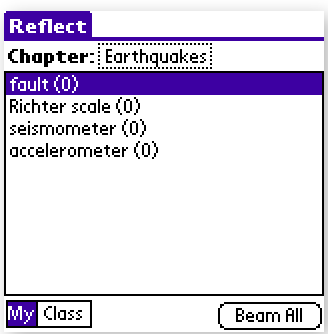

Figure 7. *Boomerang's* Prompt Screen Showing Chapters and Topics

Students Write and Categorize Questions. Using graffiti or a detachable keyboard, students compose questions in response to specific subtopics. Students can also categorize each question using a rubric devised for the purpose of the activity, which might be displayed on a board in the front of the room (Figure 8).

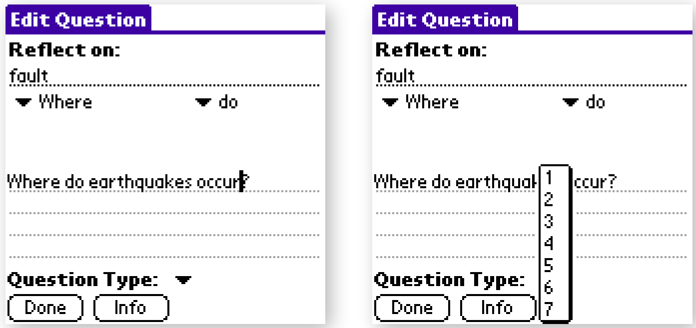

Figure 9. *Boomerang's* Question Writing Screen

Students Share Questions by Beaming. Students then share their questions by beaming to peers or to a collector device that combines the questions into a single database (Figure 9).

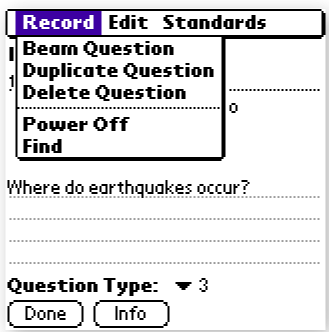

Figure 9. *Boomerang's* Beaming Function

Teacher Displays Class List of Questions. Once the questions have been collected on a single device, they can be re-beamed to all students, or the teacher can use a document camera to display the list. The list of questions can be displayed anonymously or with students' names (Figure 10).

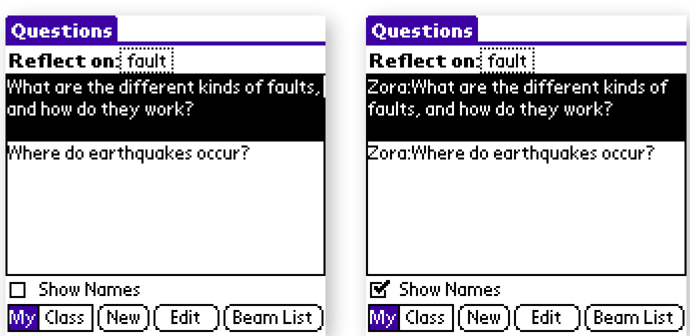

Figure 10. *Boomerang's* Displays for Class Lists of Questions

Sketchy

Many state science standards call for students to develop understanding of complex processes within biological, chemical, and physical systems. Many of these same standards require students to be able to visualize and represent those processes in diagrams, charts, and drawings. At present, very few tools exist to help teachers assess how students' understanding of complex processes through students' own visualizations of phenomena.

Sketchy is a tool that allows users to construct animations of complex scientific processes. It functions like other drawing and animation software designed for desktop computers: students use a palette of colors and tools to construct a sequence of images over multiple pages, which can then be "run" as an animation at different speeds. Software developers at the Center for Highly Interactive Computing at the University of Michigan developed *Sketchy* as part of another research project; the Image Makers team of Project WHIRL enhanced the software and developed additional classroom activities to support assessment with *Sketchy*. The team added both color and the ability to insert background images to the software. Teachers also developed and applied rubrics to judge the quality of understanding represented in student animations; in many cases, students used these rubrics to analyze and revise their own animations as they worked on them.

The flow of a typical *Sketchy* session is as follows:

Teacher Assigns an Animation for Students to Build. The classroom teacher in *Sketchy* develops a task and instructions for students to follow when constructing an animation that represents a particular scientific process. Ideally, the task assignment describes the elements students must include in their animations and includes a checklist or rubric for students to use in judging their animations as they work on them.

Students Construct Storyboards. Before building their animations, students use pencil and paper drawings to construct a storyboard of what they plan to animate. The storyboard helps students form the events or processes to be depicted in a sequence of activities (Figure 11).

Figure 11. A Student Storyboard

Students Build Animations. Individual students use the palette of drawing tools to construct individual frames, following the ideas sketched in their storyboards (Figure 12). The frames are sequenced into an animation.

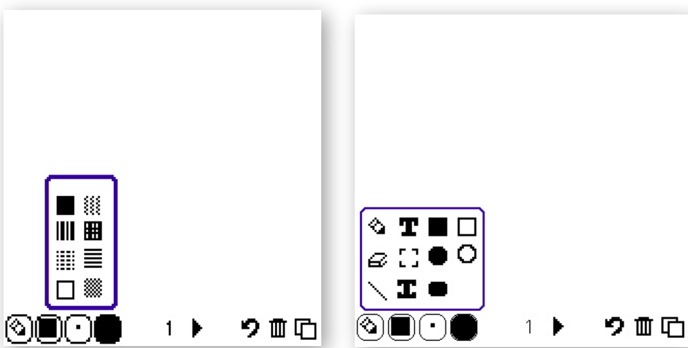

Figure 12. Sketchy's Palette of Drawing Tools

Students Review and Revise Animations. Students often take multiple class periods to build and review their animations. Students also use feedback from their peers and students in a *Sketchy* session to revise their animations.

Students Share Animations with Class. In Project WHIRL classrooms, each teacher was given a document camera to project student work for others to see. Students using *Sketchy* can put their handheld device under the camera and use it to help show their animations to others and explain the process they have depicted (Figure 13).

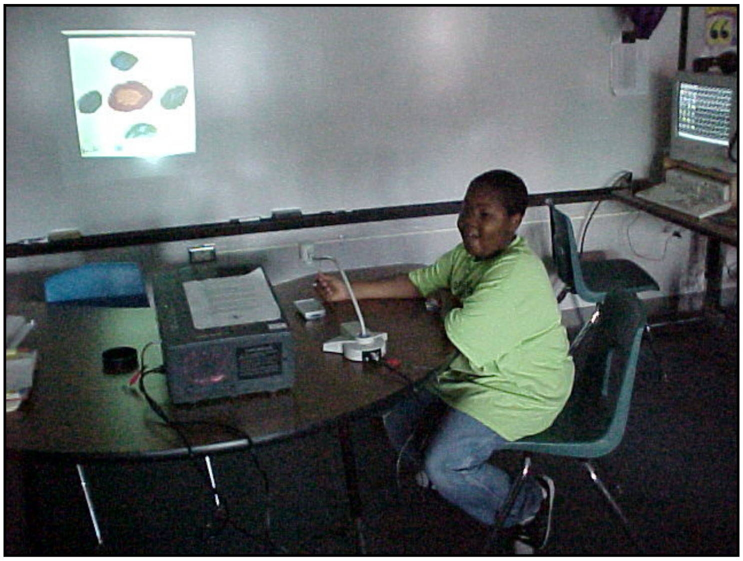

Figure 13. A Student Projects and Explains His Animation

Quizzler/Gradebook

Most all teachers currently administer quizzes and tests to assess student learning in science. Often, however, the time required to grade tests means that students do not get immediate feedback on their performance or on what they know until 1–2 days have passed. With handheld computers, it is possible for teachers to shorten the time it takes for students to receive feedback from traditional quizzes. Teachers can also use handheld-based quiz programs to help students review their understanding of concepts in the middle of the lesson or gauge students' level of understanding of concepts before a lesson begins. Teachers on Project WHIRL's design teams wanted such functionality, and we were able to identify existing tools to support this kind of assessment with handhelds.

Quizzler is a quiz application for the handheld developed by Pocket Mobility, Inc. The software allows teachers to create quizzes in a multiple-choice or true-false format on any number of topics. These quizzes can be saved and beamed to students for them to complete. All data from the quizzes can be imported into a related Pocket Mobility product that Project WHIRL teachers saw the need for and helped design, *Gradebook*. When using *Quizzler*, teachers can choose from a variety of quiz options. She can alter the sequence of questions, vary question format, allow students to repeat questions they miss in review mode, or set up the quiz as a test or exam. Teachers can also have students take a timed quiz.

The flow of a typical *Quizzler/Gradebook* session is as follows:

Teacher authors a quiz on a computer. The teacher uses a computer to construct a quiz or test using *Quizzler Maker* program. Quizzes may contain multiple-choice or true-false questions student will answer. The teacher must enter both correct and incorrect answers for multiple-choice questions (Figure 14). Project WHIRL teachers had access to research studies on typical student misconceptions in science, and they were encouraged to design incorrect answers that corresponded to typical misconceptions.

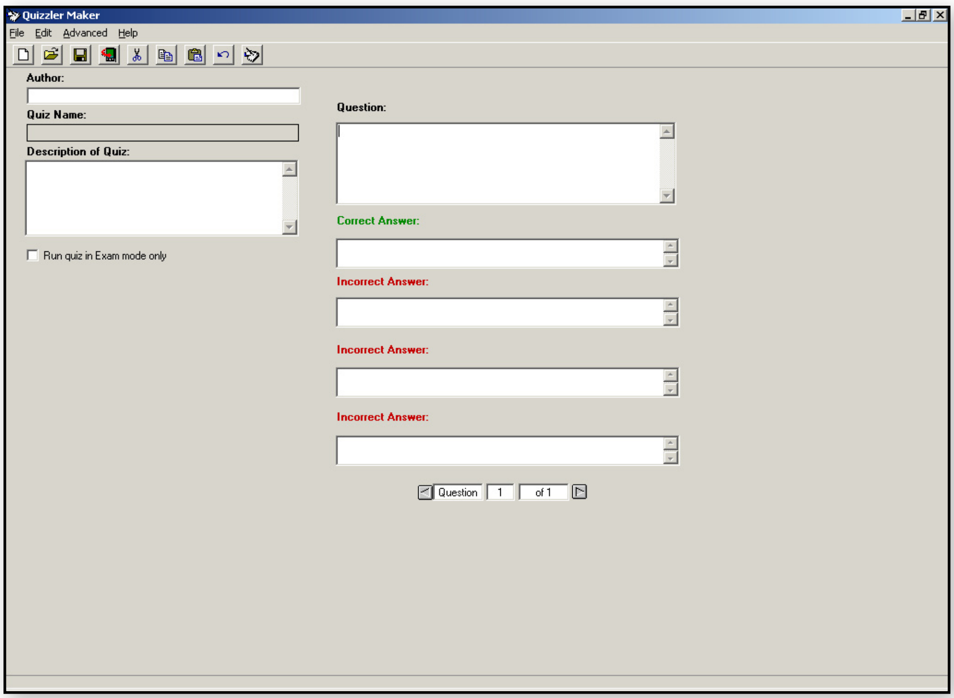

Figure 14. *Quizzler* Maker Screen Showing Answer Choices

Teacher downloads quiz to handheld and beams to students. The teacher installs the quiz to *Quizzler* on the handheld by selecting an option within *Quizzler Maker* and then hot syncs her device to her desktop. The blank quiz is beamed to all the students in the class.

Students take quiz. Students select the appropriate quiz from the list of quizzes on their handheld and tap "begin" to begin answering questions (Figure 15). The quiz can be taken in two modes. In the repeat wrong answers mode, students will get immediate feedback on their answers and have the opportunity to re-answer the question until they get it right. In the other mode, student responses are recorded and they are not informed whether or not their answers are correct.

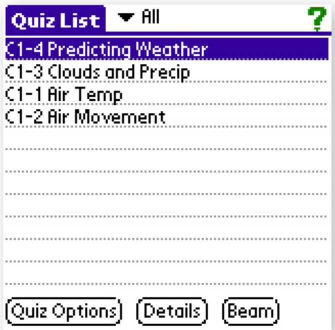

Figure 15. *Quizzler* Screen Showing List of Quizzes

Students save quiz. Students are prompted to save their scores at the end of the quiz (Figure 17).

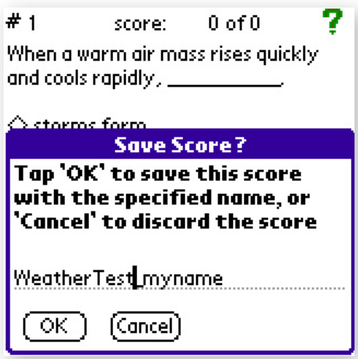

Figure 17. *Quizzler* Prompt to Save Scores on Quiz

Students review quiz results on their handhelds. Students can review results of their quiz and check questions they got wrong immediately after they have completed their test.

Teacher collects student quiz results and imports to *Gradebook*. Before Project WHIRL, individual student results were available only on the handheld. Now, students can beam their individual quiz results to a collector handheld (such as the teacher's handheld) and the quiz list file can be exported into the *Gradebook* application. Within *Gradebook*, class scores can be reviewed by question with a raw score and a percent correct score for each question. There is also a graphical display of percent correct scores.

From Process to Product: How Boomerang Came to Be

The HOT-Q team was chartered at the summer design conference with two teachers, a software developer from SRI, and the second 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 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 for how resources would be allocated to software development remained with SRI. This team ultimately struck a compromise that met teachers' current needs while 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 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 regarding concept review from observations conducted in their classrooms. For example, Sarah, the fourth grade teacher, used frequent questions posed by students 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 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 using 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 good questions.

Lessons Learned from the Design Phase

During this phase of work, we made several discoveries about the kinds of handheld software designs that teachers were likely to adopt. We 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 regarding how much new technology can import reform practice into a classroom. If we had sought to introduce a whole new pedagogy to teachers and handheld software all at once, it is likely we would have failed, especially with the HOT-Q team. Instead, the teams appeared to succeed because SRI researchers sought to develop software that could provide a focused addition 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 examples of such focused addition. 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.

In some cases, teachers discovered creative possibilities with the software that we had not anticipated. For example, in Sarah's classroom, we observed her using student questions generated in *Boomerang* to discover aspects of student understanding she had not been detecting through other means of assessing student learning. For instance, on a quiz her students failed to give the textbook answer for the definition "mixture" *t*, namely that it is a substance consisting of two or more substances mixed together but with no chemical bonding. But when the same students constructed questions, they used this definition in developing questions and the teacher felt that they demonstrated a better understanding of the concept. She also noticed that the students found the task of authoring questions motivating: "They work really hard to give you a good question. They really want theirs to be really good. Some will get picked to be on the test." Another teacher had her students create an animation in *Sketchy* showing the phases of the moon, with all the parts labeled. Her students poured over their science texts in constructing the animation, taking care to label the phases and put them in order. She observed that the attention students paid to the detail of their texts was unusual. Classroom observations from the previous year conducted in her class confirm her point of view: most students were seen to use the text only to "look for answers" and rarely as a resource from which to glean an understanding of a concept or process. On a chapter test, the teacher was surprised to find that in contrast to her other classes and classes in previous years, students got all the questions right on the phases of the moon. For her, the use of *Sketchy* proved to be useful for students in developing their ability to develop more accurate and coherent accounts of the phases of the moon.

There were also many glitches encountered in testing the software, which often led to insights into how better to support teachers in using the software. For the Data Doers team, one teacher's use of a prototype version of the software for a density lab was pivotal in helping make visible the potential benefits of *Data Doers*. As students collected measurements for mass and volume of unknown samples, the information entered into the handheld data form frequently yielded a warning; either “data out of range” or “problem with units.” As a result of those warnings, the students reconsidered their process. They frequently re-measured the sample. What could have done wrong? In several instances the same warning resulted again after re-measurement, and students measured the sample a third time. Throughout the lab activity, this teacher had students ask her for assistance if they got repeated warnings. Upon assisting teams, the teacher determined that the majority of range errors were a result of too narrow a range authored by her. She had calculated the warnings using a different kind of scale than the students were using. She shared that information with students, who were able to reflect viscerally on the importance of instrumentation in science. The students determined that they had, indeed, read the scales correctly, and therefore their values were appropriate to continue. As a result of the teacher’s discussions with students she determined that in this case the majority of errors were teacher-authoring errors, not student input errors. From this session, we learned of the need to work more closely with teachers before labs to help them think about the plausible ranges of data that students could generate within any lab. It was also discovered through this glitch that teacher errors could also achieve one of the desired effects of *Data Doers:* fostering a greater focus among students on accuracy in data collection and a willingness to re-measure values.

Finally, during the design phase we observed some uses of the tools that seemed to undercut the goal of promoting greater roles for students in reflection and self-assessment. In some cases, we observed teachers using the prompts in *Boomerang* to construct questions for students to answer, transforming *Boomerang* into a version of *Quizzler* (albeit with fewer features than Quizzler). In some *Sketchy* classrooms, we observed students paying much more attention to the colors in their animations than to accurate representations of processes. In some of these same classrooms, we observed little discussion about the processes being represented, as students worked silently on their own to build animations that looked more like jagged sequences of disconnected drawings than a true animation.

While designing the software with teachers, we decided to pay 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. Because we were concerned primarily during this phase of work with ensuring that we developed tools that met a felt need of teachers and that could be used in classrooms, we decided to worry less about initial awkward uses of the WHIRL software that seemed to undercut our own goals for improving assessment. Still, we took note of these uses and decided 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 new teachers who would participate in a field trial during the 2003–04 school year.

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

The field trial provided teachers with an opportunity to further test the software in their classrooms, develop strategies for teacher professional development, and study the effects on classroom practice and conditions for effective use of the software. In this section, we describe the field trial process, supports we provided for teachers, and key research methods and findings. In addition, we present two contrasting cases of software use, one successful and one not so successful, designed to illustrate some of the conditions for successful use we uncovered in the data from the field trial.

The Field Trial Process

The field trial began with teachers' applying to become part of the project in spring 2004. SRI staff reviewed applications from additional teachers. Once teachers were accepted, they were invited to a spring workshop planned jointly by SRI, the district, and design team teachers to introduce them to the software. Over the course of the year, teachers participated in a number of professional development activities described in the next section. Meanwhile, SRI researchers collected data from teachers on how they were using the software and on problems with the software. Teachers relied on a local professional development support teacher to relay information about problems with the software or reported them via the Project WHIRL Web site directly to SRI.

SRI software developers made minor revisions to the design of the software applications during the year to address problems teachers encountered. No major refinements to the software were required, except for *Data Doers*. During the winter, a major redesign of *Data Doers* was undertaken to simplify the flow of a *Data Doers* session and make the distribution of blank labs to students easier. Initial field trials uncovered major problems with distributing labs in the classroom; older versions were destroyed and students lost data easily with the first major release of the software. These difficulties were eliminated with the new version released during winter 2003–04.

Meanwhile, SRI researchers were engaged in research on how teachers were using the software and on the roles students were playing in assessment when teachers used the software. The focus of this research was on understanding the variation in teachers' use of the software for instructional and assessment purposes and on analyzing the conditions for effective assessment use of the software. Although we hoped to discover that teachers were using the software to discover what students know and can do and were adjusting their instruction on the basis of what they learned, we knew from previous experience to expect some teachers would not use the software in this way. By learning about the teachers and studying how and how much they participated in the professional development program we provided, we hoped to learn about the conditions that could support (and might be designed to support) effective use of the software in the future.

Professional Development and Other 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 teachers participated in 3 workshops that prepared them to use all of the project's software applications and incorporate pedagogical strategies into their practice, 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. We describe these in detail below, because they are an important part of the field trial.

The focus of the first teacher workshop was on *sample assessment activities* that were designed to introduce teachers to the software in the context of an extended investigation in science. We chose to develop materials to support activities within a Science and Technology for Children® (STC) unit, *Experiments with Plants,* which was used by a number of teachers in the project. Both elementary school and middle school teachers had familiarity with this unit, which engages students in learning how to set up a controlled experiment to investigate factors that would affect the growth of Wisconsin FastPlants™. We selected activities from the teachers guide to build assessment activities that would illustrate how WHIRL software could be used to find out what students know and can do:

- Workshop participants used *Boomerang* to develop questions at the outset of the unit and then determined whether the questions could be tested through classroom-based investigations;

- Participants collaborated in small groups to create an animation in *Sketchy* to illustrate the process of pollination as described in a text-based description in the teacher's guide; and
- Participants checked students' recorded data from selected plants using *Data Doers* and drew inferences from the datasets about which plants were "treatment" plants and which plants were "control" plants.

During the second workshop, we focused on helping teachers plan to integrate the technology into their assessment activities. We began by reviewing the activities from the previous workshop, both to remind teachers how the software works and to engage them in reflection on how the activities might be adapted and orchestrated within their own classrooms. We also introduced the idea of "backward design" (Wiggins & McTighe, 1998) to teachers, and gave them opportunities to use the state standards and their own curricular frameworks to develop assessment activities with Project WHIRL tools that would enable them to measure progress toward standards. In addition, we provided additional background material on formative assessment and gave teachers a chance to discuss principles of effective formative assessment as outlined by Black and Wiliam (Black & Wiliam, 1998b).

In the third and final workshop, teachers' sharing and discussion of how they were using the tools served as the principal focus of activities. Teachers began by sharing successful uses of the Project WHIRL software with colleagues. Next, they discussed how these functioned to support what they came to call "informal" assessment in their classrooms. As part of the workshop, teachers developed rubrics for guiding their own use of the Project WHIRL tools and worked together to plan future uses of the Project WHIRL tools. Finally, selected field trial teachers presented their approaches to integrating classroom use of the handhelds.

Each of the workshops offered significant roles for design team teachers. Design team teachers co-led hands-on activities with the software they helped to develop. They also led sessions in which they described activities that involved the use of the software in their classrooms and presented examples of student work from their class. For example, a teacher from the *Sketchy* team shared the process by which she developed and refined a rubric for use with an activity focused on the life cycle of the mealworm. A *Data Doers* team member shared his experience using *Data Doers* in a recent physics lab, in which he reported that his students' ability to see other groups' data helped motivate them to correct mistakes they discovered. All of the design team teachers also presented workshops at the National Science Teachers Association Convention in Atlanta in spring 2004. Another presented at the 2004 National School Boards Association $T+L^2$ conference.

In addition to these workshop activities, we designed a set of *curriculum planning support* materials to help teachers identify opportunities within their existing curriculum where WHIRL tools might be used in assessment activities. These include ideas for topics covered in South Carolina content standards where complex processes might be drawn and animated using *Sketchy*, as well as alignment with inquiry standards that could be accomplished by having students generate questions in *Boomerang* or check the accuracy of their data in *Data Doers*.

Once field trial versions of the software had been developed, we also developed *technical support materials* for teachers. These technical support materials included simple "feature sheets" to remind teachers of the functionality of the software. They also included more elaborated user guides that describe (with words and images) how to perform different functions of the software. Finally, we provided "Quick Guides" designed to fit on 1–2 pages that could be used by a teacher in the classroom if they got stuck.

During the field trial, all teachers in Project WHIRL were also given the option of receiving a stipend for their participation or taking part in a graduate course offered through the University of South Carolina. Graduate course credit gives teachers an opportunity to receive continuing education credits or make progress toward a degree. The course, *EDTE 671 – Computers in Science Education*, is usually offered on the Columbia campus. The course content was adapted for Project WHIRL participants and was taught jointly by the first author and district Instructional Technology Specialist Cyndi Pride. For teachers, the course provided an opportunity to explore more deeply research-based principles for designing effective formative assessment activities. The course syllabus included key readings in the field, including the National Research Council publication, *Classroom Assessment and the National Science Education Standards* (National Research Council, 2001) and Black and Wiliam's "Inside the Black Box: Raising Standards Through Classroom Assessment" (Black & Wiliam, 1998b). As part of the course, teachers were engaged in discussions about the role of student questioning in science, problems students face in collecting and analyzing data, and features of good rubrics. The culmination of the course was a presentation and write-up of an assessment activity teachers have conducted with their students as part of the project.

Local support was a critical component of our professional development design as well. A local learning technology coordinator conducted installation of the software and helped teachers set up handheld computers in teachers' classrooms. She also helped with troubleshooting problems that arose with the software and hardware. In some cases, she also provided teaching support in the classroom, modeling assessment activities or assisting teachers in leading an activity. SRI also hired a former science coordinator for the district and adjunct faculty member at the University of South Carolina-Beaufort, to serve as a *content advisor* to teachers in the project. The content advisor was a scientist and an educator who commented on teachers' designs for assessment activities and identified resources for teachers that helped teachers identify ways to incorporate WHIRL software into their classrooms. Finally, we designed the project in such a way that each school had more than one teacher from the project involved. We specifically encouraged teachers during the workshops to collaborate with their colleagues at the school and encouraged them to develop curricular and assessment materials together. This informal "buddy system" has been used successfully elsewhere to support teachers learning with technology (Penuel, Means, & Simkins, 2000).

Research 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 provided. To test our hypothesis, we designed a multi-method analytic study that featured:

- pre- and post- interview and survey measures of teachers' and students' attitudes, beliefs, and practices for monitoring progress, learning, and understanding;
- monthly classroom observations which included 64 different sessions when teachers were using the software;
- 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 are still in process. However, we have begun to synthesize data on the conditions that have been influential in supporting teachers' adoption of the tools in their classrooms. These data are summarized briefly below; in the following section we explore how these findings are illustrated by analyzing the experience of two teachers with different levels of success in adopting the tools developed or enhanced by the HOT-Q team.

Our early analyses have focused on the extent to which overall usage was associated with participation in the design process and perceptions of the tools' usability and value. For these analyses, we relied on log data as a measure of usage, group membership (design team teacher versus field trial only teacher), and survey items that focused on teachers' perceptions of the software. Those survey items asked teachers to rate on a scale from 0-3 each software application with respect to how much they liked the software, how confident they felt using it, and how much effort was required to use it.

Two factors one might have expected to predict patterns in usage did not prove to be significant. First, we did not find that participants in the design process used the software at higher rates than new field trial teachers (see Table 2 below). In fact, average number of days use showed a trend toward slightly lower use among design team teachers.

Table 2: Comparison of Design Team and Field Trial Teachers' Use of the Software (Mean Number of Days Used)

		Design Team Teachers ($n = 6$)	Field Trial Teachers (n=12)	t (df)	p
Data Doers	*M*	10.00	10.10	-.03 (14)	.98
	SD	2.83	8.01		
Boomerang	*M*	10.67	13.60	-.54 (14)	.60
	SD	8.55	11.56		
Sketchy	*M*	12.83	24.50	-1.01 (14)	.33
	SD	18.50	24.36		
Quizzler	*M*	16.67	13.60	.40 (14)	.70
	SD	16.90	13.92		

Second, we did not find that perceptions of the tools' usability or value were predictors of teachers' levels of adoption with the software. Their ratings of the software all followed a similar pattern (see Figure 17). However, teachers' perceptions of liking and of effort required to use the software were not correlated ($-.50 < r < .29$; $.08 < p < .76$) with actual use of the individual software applications, except for Data Doers, for which we found a significant correlation between liking and use ($r = .41$, $p = .04$). The co-design process may have assured that all teachers valued the functionality that the applications provided them.

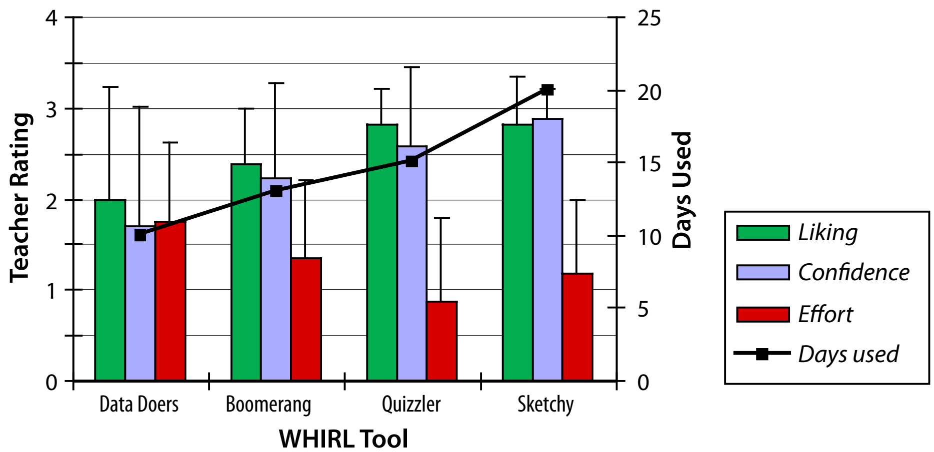

Figure 18. Pattern of Usage in Comparison to Teacher Perceptions of Use

We did find two factors that were important in helping to explain variation in teachers' adoption of the tools. One critical factor appears to be the degree to which teachers reached out either to colleagues in their school or others in the project team to get ideas about how to integrate the software into their instruction. A common characteristic of the teachers who used the software most often was that they either borrowed extensively from ideas for use shared in the workshops or relied on others in their school who were enthusiastic members of the project for ideas on how to use the software. By contrast, teachers who used the software less frequently either shared ideas with other teachers in their school less often or preferred to use a different cluster of software applications than other teachers in their school.

A second critical factor evident from the classroom observations was that teachers who seemed to find the tools easier to adopt in practice were those who were better able to manage multiple activities in the classroom while keeping students engaged. We observed that where classroom management problems were evident, teachers had many more difficulties in orchestrating use of the tools than in classrooms where classroom activities flowed more easily. We had hoped at the outset of our project that the Project WHIRL tools would not disrupt the flow of classroom activities but rather augment it; however, where flow was already an issue, the integration of handheld tools seemed to worsen the problem.

Contrasting Cases of Use: Wendy and Pamela

Two case studies of Project WHIRL teachers illustrate how these factors combined to shape different software adoption patterns. 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 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. She used the tools for a total of 19 days during the year, making the most use of *Quizzler* and *Sketchy*. By contrast, Pamela, a fourth grade teacher in the elementary school where Shawn and Sarah teach, has made use of all four of the project WHIRL software applications. She had success integrating the tools throughout her curriculum and orchestrating activities with her students. It should be noted, however, that while she made frequent use of the tools, her level of use was near the median for all WHIRL teachers, a total of 38 days during the school year.

The extent to which teachers took advantage of the professional development opportunities offered through Project WHIRL differed. Wendy attended the three required workshops, and called on the local learning technology coordinator for technical support during the field trial from time to time. By contrast, Pamela participated in the workshops and took a more active role in them. As an example, during one workshop she presented her methods for orchestrating handheld use in the classroom. Pamela also called on the local learning technology coordinator more often than Wendy for both technical and pedagogical support. Pamela 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 played a more central role within it.

In addition, despite the fact that both teachers joined the project with other teachers in their school, they relied on their colleagues to different degrees. Pamela approached her colleagues often for help, and sought out advice from SRI staff that could benefit not just her but her other colleagues in the school. She and her colleagues shared ideas for how to use the software on a regular basis. By contrast, Wendy did not share ideas with her colleagues. The three teachers in that school had divergent teaching styles and also believed that they served groups of students that had different needs from one another.

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, when 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 uses *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 in 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.

Discussion

In recent years, software development efforts undertaken by researchers in the learning sciences have produced a number of powerful, high-quality technology innovations to support science learning (Roschelle, Pea, Hoadley, Gordin, & Means, 2000). For example, researchers have been successful in developing innovations that use technology to teach key concepts in Earth Science (Crawford & Toyama, 2002; Feldman, Konold, & Coulter, 1999), biology (Reiser et al., 2001), chemistry (Schank & Kozma, 2002), and physics (White & Frederiksen, 1998). More recently, learning scientists have also begun to develop a wide range of powerful applications to support science learning that rely on handheld computers (Soloway et al., 1999; Tinker & Krajcik, 2001).

Many of these educational technology innovations have proven effective in a limited number of classroom contexts; however, far fewer have been successful in scaling up (Edelson, Gordin, & Pea, 1999). A chief obstacle has been that most innovations, while of high quality, are not usable by a broad range of teachers without extensive support from researchers (Blumenfeld, Fishman, Krajcik, Marx, & Soloway, 2000; Fishman & Krajcik, 2003). Further, many innovations are undercut by the existing capabilities of schools and districts and a culture within schools that limits opportunities for collegial sharing and critique of ideas for teaching.

One of the more promising aspects of handheld computers which recent studies have underscored is that software applications that run on handheld computers tend to be more easily adopted by teachers than desktop-based applications. SRI's evaluation of the PEP program found that the tools were readily integrated into students' learning activities (Vahey & Crawford, 2002). Similarly, researchers in the Center for Highly Interactive Computing in Education at the University of Michigan indicate that teacher adoption rates of handheld software the Center has developed are nearly twice as high as adoption rates of their Web-based tools (Fishman, personal communication). In Project WHIRL, teachers used at least one of the software applications an average of 41 days out of the school year. Our high overall rates of adoption in Project WHIRL are consistent with this finding, and were likely aided by our specific attention to creating usable software through a design process that involved teachers throughout the development process.

Nonetheless, our case studies illustrate that handheld computers are not adopted at similar levels by all teachers who are given a classroom set. Teachers' reliance on their ties to other teachers in the school and on professional development supports helped to explain, in part, some of the variation we saw in rates of adoption. In addition, teachers' own skill in managing their classrooms shaped their use. For some teachers, introducing handheld computers exacerbated an already difficult situation in which the teacher felt ill-equipped to orchestrate multiple activities within the science classroom or lab. For others, incorporating handhelds was a natural extension of their management of instruction.

There are also some costs to consider in creating tools that are usable by a wide variety of teachers. It may be that tools such as ours do too little to support inquiry-oriented science teaching practice in the ways that earlier software developed for desktop computers did. They may in fact be easily used to support the kind of practice we observed in some cases, where a tool like *Boomerang* designed to support student questioning is used to support further teacher questioning of students through the use of questions similar to those found on multiple-choice tests. Tools like *Quizzler,* which map closely onto traditional assessment practices, certainly reinforce the notion that multiple-choice quizzes – which can put students in more passive roles with respect to developing deep subject matter understandings – are the only suitable form of assessment. SRI's representatives on the design team had to work hard to convince teachers o even try *Boomerang*; if there had been no attempt to influence this team's direction, the teachers might have been satisfied to use Quizzler alone to support classroom assessment.

Other design processes privilege subject matter expertise more than co-design does, which may reduce the risks associated with taking teachers' ideas as the basis for software design. Research efforts in the learning sciences, for example, have often drawn more from experts in science or in psychology to support the development of their technologies (see, for example, Parr, Jones, & Songer, 2002). Both groups of experts bring to the design process specialized knowledge about the structure of subject matter content and about opportunities to enhance student learning opportunities, knowledge that many teachers may not feel that they have. However, without that knowledge, teachers who implement technology-supported curriculum materials developed by experts may do so superficially and fail to incorporate significant aspects of the designs into their teaching (Brown & Edelson, 1998).

Developers of handheld software applications designed to support assessment are then left with what is an irreducible tension between developing tools that meet teachers' perceived immediate needs and developing tools that support improvements in teaching and learning desired by reformers in academe and government. Our own experience and research in Project WHIRL suggests that co-design in which researchers' expertise informs the design and testing of software alongside teachers' wisdom of practice is a powerful way to produce more usable technology innovations for education. In addition, designing collegial and professional development supports for teachers is critical for establishing conditions that support adoption in ways that truly extend teachers' range and quality of assessment practice.

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*.

Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist, 35*(3), 149–164.

Brown, M., & Edelson, D. C. (1998). Software in context: Designing for students, teachers, and classroom enactment. In A. S. Bruckman, M. Guzdial, J. L. Kolodner & A. Ram (Eds.), *Proceedings of ICLS 98:International Conference on the Learning Sciences* (pp. 63–69). Charlottesville, VA: AACE.

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

Crawford, V. & Toyama, Y. (2002). *WorldWatcher Looking At The Environment Curriculum: Final external evaluation report*. Menlo Park, CA: SRI International.

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

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *The Journal of the Learning Sciences, 8*(3 & 4), 391–450.

Feldman, A., Konold, C., & Coulter, B. (1999). Network science, a decade later. *Hands On!, 22*(2), 1–2, 16–18.

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.

Fishman, B. J. & Krajcik, J. (2003). What does it mean to create sustainable science curriculum innovations? A commentary. *Science Education, 87*(4), 564–573.

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.

Ogle, D. S. (1986). K-W-L group instructional strategy. In A. Palincsar, D. S. Ogle, B. F. Jones & E. G. Carr (Eds.), *Teaching reading as thinking (Teleconference Resource Guide)* (pp. 11–17). Alexandria, VA: Association for Supervision and Curriculum Development.

Parr, C., Jones, T., & Songer, N. B. (2002). CyberTracker in BioKIDS: Customization of a PDA-based scientific data collection application for inquiry learning. In P. Bell, R. J. Stevens & T. Satwicz (Eds.), *Keeping learning complex: The Proceedings of the Fifth International Conference of the Learning Sciences (ICLS)* (pp. 574–575). Mahwah, NJ: Erlbaum.

Penuel, W. R., Means, B., & Simkins, M. (2000). The multimedia challenge. *Educational Leadership, 58*, 34–38.

Penuel, W. R., Roschelle, J., & Shechtman, N. (in preparation). *The WHIRL Co-Design Process: Participant experiences*.Unpublished manuscript, Menlo Park, CA.

Penuel, W. R., Tatar, D., & Roschelle, J. (2004). The role of research on contexts of teaching practice in informing the design of handheld learning technologies. *Journal of Educational Computing Research, 30*(4), 331–348.

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

Reiser, B. J., Tabak, I., Sandoval, W. A., Smith, B. K., Steinmuller, F., & Leone, A. J. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263–305). Mahwah, NJ: Lawrence Erlbaum.

Roschelle, J., Pea, R. D., Hoadley, C. M., Gordin, D. G., & Means, B. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children, 10*(2), 76–101.

Schank, P. & Kozma, R. B. (2002). Learning chemistry through the use of a representation-based knowledge building environment. *Journal of Computers in Mathematics and Science Teaching, 21*(3), 253–279.

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), 30–37.

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

Vahey, P. & Crawford, V. (2002). *Palm Education Pioneers program: Final evaluation*. Menlo Park, CA: SRI International.

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

Wiggins, G. & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.

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.

Author Biographies

William R. Penuel is senior educational researcher at the Center for Technology in Learning at SRI International. His research focuses on assessment and evaluation of technology-supported initiatives in science and mathematics in both school and community-based settings. He served as the Principal Investigator for Project WHIRL.

Louise Yarnall is an educational researcher at the Center for Technology in Learning at SRI International. Her research focuses on evaluation of technology-supported initiatives in school settings and assessment of student learning from media-based materials. She served as Co-Principal Investigator for Project WHIRL.

J·T·L·A

The Journal of Technology, Learning, and Assessment

www.jtla.org

Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College