HANDHELD TOOLS THAT "INFORMATE" ASSESSMENT OF STUDENT LEARNING IN SCIENCE: A REQUIREMENTS ANALYSIS

Jeremy Roschelle
William R. Penuel
Louise Yarnall
Nicole Shechtman
SRI International

Deborah Tatar
Virginia Tech

Contact

Jeremy Roschelle

Center for Technology in Learning

SRI International

333 Ravenswood Ave.

Menlo Park, CA 94025

Jeremy.Roschelle@sri.com

Phone: + 1 650 859-3049

Fax: + 1 650 859-4605

In Press, Journal of Computer Assisted Learning.

Published in the Proceedings of the 2nd International Workshop on Mobile and Wireless Technologies in Education, JungLi, Taiwan, March, 2004.

Voted Best Paper.

HANDHELD TOOLS THAT "INFORMATE" ASSESSMENT OF STUDENT LEARNING IN SCIENCE: A REQUIREMENTS ANALYSIS

Abstract

An important challenge faced by many teachers as they involve students in science investigations is measuring ("assessing") students' progress. Our detailed requirements analysis in a particular school district led to the idea that what teachers need most are ways to increase the quality of the information they have about what students know and can do, not automation of typical assessment practices. We see handheld computers as promising tools for addressing this need because they can give students and teachers frequent, integral access to new ways of expressing and communicating what they know and can do. Our requirements analysis has led us to emphasize a need for handheld-based tools that "informate" science instruction by:

- Being oriented to the needs of teachers in transition to inquiry-oriented pedagogy;
- expanding the range of assessment tasks through a new representational medium and communication infrastructure;
- creating new roles for students in expressing what they know and can do; and
- focusing both students' and teachers' attention on scientific concepts.

Keywords

Science learning, Assessment, Requirements Analysis, Classroom Research

HANDHELD TOOLS THAT “INFORMATE” ASSESSMENT OF STUDENT LEARNING IN SCIENCE: A REQUIREMENTS ANALYSIS

Introduction

Teachers have found handheld computers useful for supporting student investigations in science classes (Vahey & Crawford, 2002). For example, in many classrooms, students have used graphing calculators or palm-sized computers adapted with probe kits to gather and graph data. Educational researchers have used handhelds in science to support bird observation and concept mapping, measure the quality of collaboration in small groups, and create participatory simulations of scientific processes, among many other applications relevant to improving science learning (Chen, Kao, Sheu, & Chiang, 2002; Kaput & Hegedus, 2002; Novack & Gleason, 2001; Solomon & Perkins, 1998; Soloway et al., 1999; Stroup, 2002; Tinker & Krajcik, 2001; Yarnall et al., 2003).

An important challenge faced by many teachers as they increasingly involve students in science investigations is measuring (or “assessing”) students’ progress. A story we collected from one elementary school teacher, whose fifth graders have won a top science fair award, illustrates this challenge. With her support, students independently designed a wonderful scientific study that involved advanced concepts of perception and human psychology: they investigated whether presenting a color with a word could enhance memory for words. The teacher reported that students’ hypotheses focused on the dual influences of literacy and color on memory. Students believed that being older and able to read better made color less important to helping students remember a word. Despite the advanced quality of the students’ work, in the end, they were graded on the basis of whether they had completed steps in the project rather than on the basis of whether they understood the content and the design of their study.

The misalignment between content and assessment of students’ work that can be observed in this story occurs in many other teaching and learning settings. Furthermore, researchers who have studied the implementation of inquiry-based approaches to science instruction have found that when teachers do not use formative assessment strategies to support instruction, students learn less (Barron et al., 1998; Chen et al., 2002; Petrosino, 1998; White & Frederiksen, 1998). Researchers have also found that when teachers use traditional summative assessments, such as tests, in addition to investigations, students can get the impression that investigations are "for fun" and not related to their teachers' expectations (Means & Haertel, 2002; Means, Penuel, & Quellmalz, 2001; Young, Haertel, Ringstaff, & Means, 1998). Teachers need support to conduct ongoing formative assessments to track students' evolving content and conceptual knowledge. As the literature review that follows indicates, improved formative assessments can have a large effect on student achievement. Yet it is often difficult for teachers to adopt new formative assessment practices.

Our Wireless Handhelds Improving Reflection on Learning (WHIRL) project (http://www.projectwhirl.org) is exploring the use of handhelds to provide support for better assessment in science classes. We have focused on a requirements analysis to define what software capabilities teachers may need. Through such analysis, we sought to identify ways in which teachers could reap the benefits from handheld formative assessments while avoiding adoption barriers.

Our work has revealed one key theme: teachers need tools that "informate" rather than "automate" assessment. The distinction comes from the work of Zuboff (1988), a sociologist who studied how new information technology (IT) affects workers. She argued that although many IT implementations result in deskilling workers, some IT implementations can actually enable workers to become more knowledgeable, effective, and empowered. These implementations create more informative displays that enable workers to learn and improve their performance on the job. Zuboff coined the term "informate" to describe this possibility:

> Information technology not only produces action but also produces a voice that symbolically renders events, objects, and processes so that they become visible, knowable, and shareable in a new way. Viewed from this interior perspective, information technology is characterized by a fundamental duality that has not yet been fully appreciated. On one hand, the technology can be applied to a logic that hardly differs from that of the nineteenth-century machine system—replace the human body with a technology that enables the same processes to be performed with more continuity and control. On the other, the same technology simultaneously generates information about the underlying productive and administrative processes through which an organization accomplishes its work. It provides a deeper level of transparency to activities that had been either partially or completely opaque. In this way, information technology supercedes the traditional logic of automation. The word that I have coined to describe this unique capacity is informate. (p. 9)

Our analysis has led us to believe that the concept of "informating" is appropriate to the problem of assessment in inquiry-oriented science instruction because what teachers need most are ways to increase the quality of the information they have about what students know and can do, not just automation of what they do already to measure students' progress. We see handhelds as promising for this purpose because they can give students frequent, integral access to new representational forms and communication options, which could enable students to better express what they know and can do.

Project WHIRL progressed from the initial requirements analysis to a participatory design process in which we worked with three teams of teachers to design or adapt handheld tools for assessment. Currently, we are in the midst of a one-year evaluation of the three resulting tools with a larger group of science teachers. Although the results are not yet in, we believe our requirements analysis bears reporting because assessment is an exceptionally important area of functionality for all handheld learning applications. The concept of assessments that "informate" has broad applicability to the wide variety of areas in which researchers are applying wireless handhelds to the problem of improving learning.

A key principle throughout our requirements analysis has been "reciprocal influence" _ we sought to define a process in which ideals from research and practical wisdom from teaching come into active interplay. Consequently, our paper begins with a presentation of the research and teaching contexts that set the stage for handheld assessment. Following this review, we introduce the contrast between handheld assessment tools that "informate" or automate. We illustrate the distinction with a description of one tool that was refined for use in science classrooms by a participatory design team.

Handheld assessments: Research context

Interest in classroom assessment derives in part from results of a wide range of research studies that suggest the value of formative assessment, that is, assessment that yields knowledge that teachers can use to adapt and improve instruction. When assessment practices enhance feedback between teachers and students, involve students in self-assessment, and support students' motivation for learning, they can lead to significant benefits for students and teachers (Black & Wiliam, 1998). The benefits to students cited by researchers for improving classroom assessment include increased self-regulation and correction of errors (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler & Winne, 1995), increased ability of students to self-assess the quality of their own work (Rudd & Gunstone, 1993), increased equity in classrooms (Cole, Coffee, & Goldman, 1999), a greater willingness on the part of students to review and revise their ideas (Cognition and Technology Group at Vanderbilt, 1997), and better performance on summative, external tests of achievement (Black & Wiliam, 1998; Crooks, 1988; Fuchs & Fuchs, 1986). Researchers also cite potential benefits to teachers: as classroom assessment yields better and more frequent data about student progress, teachers may be motivated to rethink their teaching practices and expectations for student learning (National Research Council, 1999, 2001). In summarizing their findings from a meta-analysis of 43 studies of classroom assessment's effects on student learning, Black and Wiliam (1998) concluded:

> There is a firm body of evidence that formative assessment is an essential feature of classroom work and that development of it can raise standards. We know of no other way of raising standards for which such a strong prima facie case can be made on the basis of evidence of such large learning gains. (p. 19)

At present, most teachers' assessment practices are not likely to lead to such large learning gains, because they do not have the characteristics of good formative assessment. Research studies indicate that teachers' assessments rarely match teachers' more ambitious instructional goals. Even when teachers engage in more student-centered instruction, they tend to use assessments that focus on recall of facts (Fleming & Chambers, 1983; Stiggins & Conklin, 1992). In many cases, teachers do not understand well how the methods they use align with the kinds of achievement they are assessing; nor do teachers understand how to sample student performances to gauge learning accurately or how to avoid bias (Stiggins & Conklin, 1992).

Research on assessment reforms shows both some promise and persistent challenges to improving assessment. A number of these reforms have been successful in achieving short-term improvements in the quality of teachers' assessments and in student learning (Black & Harrison, 2001; Shepard, 1997). At the same time, reformers have discovered that teachers rarely have adequate time to plan assessment activities in a principled manner or to learn new strategies for assessment from peers and experts (Darling-Hammond, Ancess, & Falk, 1995). When teachers do succeed in collecting more varied forms of data on student assessment, they often experience "information overload" (Black & Wiliam, 1998; National Research Council, 2001). Some teachers have experienced their involvement in assessment reform efforts as threatening the core of their professional identities and as demanding fundamental rethinking of strategies for solving the day-to-day dilemmas of teaching (Atkin & Black, 2003; Atkin, Sato, Coffey, Moorthy, & Thibeault, 2003).

Handheld assessment: Exploring the classroom context

Sources of data

We based our requirements analysis on detailed interviews and observations within a particular U.S. school district. The Beaufort County, South Carolina, district was selected because of its prior experience in implementing laptops and Palm computers and the broad diversity of students and economic conditions present there.

Interviews took place in the spring of 2002. A structured interview protocol was designed to elicit teachers' perspectives on the core issues and concerns related to project-based science instruction and assessment. It focused on both teachers' background and their experiences in conducting and assessing project-based science lessons. The protocol included detailed questions about learning goals, project activities, and the degree to which students were allowed to direct aspects of their own projects. It also asked about formal and informal ways teachers assessed student knowledge before, during, and after projects. Each interview was conducted by two researchers and took about an hour.

We sought to interview as many teachers as possible who taught science in grades 4 through 9. Twenty-five teachers were identified by the district's technology coordinator and by other interviewees. The sample was diverse with respect to teaching assignment, gender, ethnicity, socioeconomic status of students, and professional experience. It included 7 elementary school teachers (3 classroom teachers and 4 science specialists), 10 middle school teachers, and 8 high school teachers. Biased toward females, the sample included 6 men and 19 women. Biased toward white teachers, the sample included 4 who were African-American and 21 who were white. One participant was from the Gullah community, a distinctive African-American cultural group that has lived in the Beaufort area for hundreds of years. The teachers were evenly distributed across three geographic clusters of the county. Location can be considered a proxy for socioeconomic status of students, since the southern part of the county is more affluent and the northern area is more working class. Teachers also had a wide range of experience levels: just under half (12) of the teachers we interviewed had been teaching 5 years or less, but nearly a quarter (5) had been teaching for more than 20 years. Some teachers had emergency credentials, and others had full certification. Some of the fully certified teachers also had master's degrees. About a third (8) of the teachers had some job-related experience working as scientists or assistants to scientists, while science had not been a particular focus of preservice education or subsequent professional development for the others.

Observers visited six of the participating teachers' classrooms. Researchers trained in ethnographic observation techniques visited individual classrooms three or four times in a single week, noting in detail the organization of classroom activities, discourse, assessment strategies, and participation structures.

Another major source of information was notes from seven teachers' participation in a co-design workshop in July 2002. The initial workshop could be considered a culminating part of the requirements analysis process, since in this workshop we learned concretely what kinds of needs for, approaches to, and benefits of the use of handhelds Beaufort teachers and community members found compelling. Activities included teambuilding, brainstorming ideas for handheld software that could support assessment in the science classroom, and creating charter documents that described the classroom need each tool would address.

Our data analysis strategies were guided largely by our overall project goals and by the timeline of our grant. Our project aimed to promote assessment of inquiry-oriented science activities with handheld computers, and we needed to complete our data analysis quickly over the course of 3 months (summer 2002) so that the analysis could inform the design process. We combined sampling and analytic techniques typically employed in grounded theory (Strauss & Corbin, 1990) to help us assess the distance between teachers' current practice and our goals and used rapid ethnographic analysis techniques (Millen, 2000) to make sense of the data quickly in time to inform the design process.

Our sampling approach was theoretically-driven, in that our focus on teachers in grades 4-10 was expected to reveal developmental differences in teachers' emphasis of particular inquiry skills as outlined in the National Science Education Standards (National Research Council, 1996). Students are expected to master new inquiry skills as they transition from elementary to middle school and then again to high school; teachers' reports of their practice were expected to reflect students' challenges at each of these transition points. To analyze teachers' reports of their practice, however, we did not rely on the standards documents but rather constructed a coding scheme for teachers' instructional and assessment strategies inductively, drawing on the language teachers themselves used to characterize practice. Only after we analyzed the data in this way did we examine the themes from the interviews in light of what is called for in the Standards. The results of those analyses are presented in summary form in the next section of this paper.

We then presented these data to our fellow researchers and designers on the project in a project-wide meeting, in order to develop some guidelines for design and establish a design process to follow. We developed a set of potential scenarios (see Millen, 2000, for a description of this analytic technique) as a group to guide us. When the design process began with teachers in late summer 2002, we continued developing scenarios with teachers. On the basis of these scenarios worked out over the course of the summer months, we developed the set of design guidelines for handheld assessments that are presented in this paper.

The classroom context in Beaufort

Teaching assignments. Teachers had varied types of teaching assignments. Some elementary teachers saw students all day long in an "intact classroom" where science was one of multiple subjects being taught. Most middle and high school teachers taught science as a single subject to multiple classes of students each day; classes in these cases met two or three times per week, but for various lengths of time (from 45 to 70 minutes per class). An especially challenging assignment that some elementary teachers faced was being a "lab coordinator" for their school; these teachers saw all K-5 students once per week in their lab classrooms, where their chief assignment was to lead hands-on activities in support of science standards. One teacher who had such an assignment illustrates the impact of teaching assignment on the ability to do extended projects and detailed assessment. She noted:

> Seeing them once a week, I can't really require research and take 500 papers and grade them. It just is not practical, and that's not really what I'm supposed to be doing anyway. I'm supposed to be providing that hands-on time that the classroom teacher has difficulty getting to.

Learning Goals. Among the teachers we interviewed, we found that teachers' goals for science instruction varied considerably across grade levels but were generally consistent with expectations of national and local standards for students' inquiry skills. Elementary school teachers reported focusing on familiarizing students with the tools of inquiry (e.g., how to use a microscope) and with different processes and systems underlying familiar phenomena (e.g., their own bodies, animals and plants around them). Middle and high school teachers said they were concerned primarily with helping students understand scientific concepts and principles. They sought to illustrate those concepts through hands-on activities and exercises in the classroom. Students at these levels in Beaufort, teachers reported, engaged in more phases of inquiry independently or within groups at higher levels, but some teachers had concerns about which members of these lab groups were doing most of the work and learning.

Use of project-based science instruction. On the basis of our initial visits to Beaufort prior to the project's inception, we anticipated that many teachers would report that they frequently used extended projects in science. The school district has a strong commitment to teaching science through inquiry, and inquiry skills are an important part of the South Carolina standards. A few schools in the district use interdisciplinary teaching team methods in which extended projects feature strongly. In practice, however, we found that project-based instruction in science was relatively infrequent.

From our initial analysis of interviews with teachers, we found that although nearly all the teachers could describe projects that they conducted with students in science, few conformed to researchers' typical notions of project-based science. Few projects lasted more than two or three class periods, and although nearly all the projects involved students' creating some product or artifact, most lacked a driving question (see Blumenfeld, Soloway, Marx, Guzdial, & Palincsar, 1991). Many teachers' classroom schedules did not permit them to engage students in extended projects; it is unclear from our research to date how much opportunity teachers have had to learn about student-led inquiry.

At the same time, most of the teachers believed in the power of hands-on science learning activities to clarify difficult science concepts, especially at the secondary level. Teachers reported (and we observed) that hands-on activities were relatively frequent. However, teachers used more traditional forms of assessment to assess those skills. Many of them did employ informal techniques of assessment to guide their judgments about what their students learned, but they had few ways to use their observations to improve their instruction.

Assessment goals. Teachers' goals for assessing student learning were multifaceted. In the interviews, teachers made many mentions of using grades from students' homework, tests, and quizzes to analyze students' factual knowledge or conceptual understanding. Several interviewees cited the accurate use of scientific vocabulary as a dimension of student performance that they thought they should assess. A number of teachers also sought to use classroom assessments to measure students' participation in terms of level of effort and engagement and the quality of students' collaboration or cooperation with their peers. Three teachers also cited a focus on having students follow directions.

Teachers were also concerned with using assessment to find out whether students' factual or conceptual knowledge was somehow "connected," and not simply isolated bits of knowledge. They hoped that they could use assessments to find out whether students could make a connection between their classroom science learning and world events. Teachers also used classroom assessment as a vehicle to encourage students to think about connections between science concepts and their own lives. Third, teachers hoped to use classroom assessment to foster the development among students of a better conceptual map of the topic or domain being studied, by encouraging students to make connections among concepts.

There were a few mentions of assessment goals that are consistent with researchers' recommendations for high-quality classroom assessment. A number of teachers mentioned that a focus of their assessment activities was on getting students themselves to reflect on what they were learning. Also, a few teachers sought to use assessment as a source of feedback on their own teaching, so that they might improve the quality of their instruction. Both of these goals map well onto Black and Wiliam's (1998) recommendations that classroom assessments involve students actively in monitoring their own learning and that assessments be used to inform instruction.

Assessment strategies. In the interviews, the most commonly cited strategies employed by teachers were formal, such as tests and quizzes, and grading of student assignments using rubrics. Teachers in the study did mention using informal strategies embedded within their instruction to assess students, especially observation and questioning. Despite the fact that a number of teachers mentioned self-reflection as a focus of assessment, only two teachers cited student self-assessment as a technique they used in their classrooms. Two teachers said they opposed formally assessing students, and two teachers said they opposed the use of any informal assessment techniques.

Technology fluency and use. Teachers varied in their level of experience with using technology. Some had very limited background in using technology. One teacher, for example, used the Internet periodically with her students, mostly to look up information or visit a Web site as part of an extension activity to a lesson. Although she had owned a personal computer for some time and used it to create worksheets for students and write to parents and faculty members, she was dependent on her husband to help troubleshoot any problems she encountered. He was "the computer person," and he was her first resource for anything technological. When we first met this teacher, she took few risks on her own with technology and was quick to attribute any difficulties with technology to her own lack of skill or knowledge.

By contrast, because of the influence of the district's prior experience with laptops and handhelds, some teachers were comfortable using technology in the classroom. One teacher used her computer throughout the day with her students, and she had experience using handhelds with her students as well. She used the handheld computers in her classroom to support some more formal forms of assessment. For her older students, she developed checklists on her own device and then beamed them to students to check off the steps they had done in an activity or lab. She had students turn in their checklists, so she could assess whether or not students had completed the assignment. These checklists also helped students keep track of their own progress, an important component of the many hands-on activities that she facilitated.

Community Culture. We also found that community culture could have a strong impact on teaching practices. From long talks with a key informant from the Gullah community, we came to understand that the introduction of handheld assessment in Beaufort would be seen as a civil rights issue by many Gullah parents. Gullah parents exert considerable influence over what kinds of new teaching practices can succeed in Beaufort schools, and it would not be easy to gain their trust. This informant emphasized that handhelds would have to be seen as empowering Gullah students to be successful: to enable the students to become more expressive and capable and to suggest tangible trajectories for them to higher-paying jobs in the future. It would be particularly beneficial if students could use the handhelds to demonstrate to their parents how they were learning more. From our observations of Gullah students in an after-school, community-based program, we suspected they could rapidly become expressive users of computer technology in project-based settings.

A unifying theme from the study: Teachers "in transition"

As we investigated various aspects of these teachers' practices, we found that a number of teachers appeared to be what we call "in transition" to inquiry-oriented science teaching. According to the National Science Education Standards (National Research Council, 2000), some key practices that are essential to inquiry-based science teaching include: (a) selecting content that meets the interests and understanding of students, (b) challenging students to take responsibility for their own learning, (c) encouraging and modeling the skills of scientific inquiry, and (d) encouraging collaboration among students. Assessment practices encouraged for inquiry-based science teaching include using multiple methods and systematically gathering data about student understanding and ability, using assessment data to guide teaching, and guiding students in self-assessment. The teachers all supported the goal of promoting inquiry in the classroom, but most of the teachers met some of the inquiry standards but not others. In most cases, although teachers used frequent hands-on, collaborative projects, students often used science kits to conduct experiments, rather than formulating their own questions and designing their own investigations. In addition, teachers' assessment practices tended to lag behind other inquiry pedagogy. Hence, we have come to the conclusion that many teachers were "in transition" to inquiry-oriented science.

A middle school teacher we interviewed and observed provides an example of the transitional instruction we found. Many of the activities in her classroom had elements of inquiry-oriented science instruction. For most activities, this teacher's students were organized into peer-led cooperative groups. She gave group members responsibility for undertaking a series of tasks and reporting back to her on their progress. She assigned different kinds of leadership roles for different units; these roles typically were given names appropriate to the topic they were studying. For example, for the unit she taught on rocks and minerals, one student in each group was the "head geologist." The teacher played a facilitative role with the groups, moving from group to group as the students worked largely on their own. She also organized groups of cooperative activities into "units" that often had a real-world connection or driving question. Her "Deltarian" unit, for example, designed to cover both earth science and biology strands in the standards, involved students in deciding whether Earth would be suitable for colonization by the alien "Deltarians."

However, many aspects of her lessons were not well aligned with teaching standards for inquiry. For example, although these units sometimes have a driving question, they were not extended student-centered projects in the way that many researchers think of these (see Blumenfeld et al., 1991). Furthermore, assignments were all laid out for students ahead of time, so students did not have to decide how to allocate resources to complete the project.

One elementary school teacher whose classes we visited provides another example of someone whose practices were "in transition." She discussed her Fast Plants activity, in which students learn about the life cycle of plants. During this unit, she engaged her students in extended projects involving the study of the conditions in which Wisconsin Fast Plants grow best. Through the course of the project, the teacher noted:

> Children are able to observe them quickly, go through the entire plant cycle. And they actually do the pollination. We start with a seed; they go all the way to the flowers; we pollinate the plants, cross-pollinate. They have the seedpods; we actually harvest those seeds, and I use those seeds again the next year with another group. And so it involves a lot of observation, a lot of collecting of data and those good techniques that are important for the elementary students to learn to do.

She provided students with a context for helping them to understand the larger significance of what they were doing in the project, too, in terms of the interconnections between animals and plants on the Earth:

> Today we were talking about the fact that not all the plants are growing at the same rate, that they would understand that there's change; and we're also going to work on environmental issues and the fact that we're all part of one planet and that all the animals and plants interact.

Despite the hands-on emphasis and conceptual grounding this teacher provided, she did not have a method for assessing the degree to which her students linked their understandings of the concepts to the hands-on activities. In the Fast Plants project, for example, she required students to observe and record observations and harvest the seeds. We would predict, on the basis of some of the research findings reviewed earlier, that it is unlikely that her students would have been able to make the link between the concept of life cycles she was hoping students would learn and the activities they undertook as part of the Fast Plants project. They implemented these experiments but perhaps did not understand how the experiments related to deeper concepts. The teacher did not have a way to collect evidence that could show whether her students made these connections.

Guidelines for handheld assessments that informate

On the basis of the requirements analysis discussed above, we formulated four key guidelines for handheld assessments that could be effective and widely adoptable in Beaufort schools.

Guideline #1: Design for teachers "in transition."

An overarching guideline was that handheld assessments should be designed for teachers "in transition" to inquiry-oriented science teaching. Many researchers hold an ivory-tower image of inquiry-oriented science teaching. In their idealized image, students engage in long-term projects, with substantial opportunity to design and carry out their own investigations. Science concepts arise within these investigations, and teachers coach and facilitate conceptual convergence. Frankly, the practices of very few teachers fit this ivory-tower image, and thus tools designed exclusively to support it are not likely to be very useful. We instead became more focused on teachers' trajectories toward adopting strategies from an inquiry-oriented approach, such as encouraging students to come up with questions and predictions, to draw diagrams and models, and to recognize more readily when their instruments are recording bad data. In this trajectory toward inquiry-oriented teaching strategies, a central emerging issue was that teachers' assessment practices often lag in the transition to an inquiry orientation. Thus, a key design guideline we settled on was that handheld assessments should help teachers initiate assessment practices that were better aligned with the purpose of the aspects of inquiry-oriented instruction they had already adopted.

In working with this guideline, we found that the word *assessment* was not particularly useful, since it often conjures up images of quite traditional tests. We therefore used the question "How can you capture more about what students know and can do?" to orient our design teams as we focused them on how handheld assessments could improve the effectiveness of those aspects of inquiry instruction that they were implementing in their classrooms.

Our requirements analysis pointed to four additional guidelines, each of which became important in our subsequent participatory design process.

Guideline #2: Exploit the unique representational capabilities of handhelds compared with paper.

In other work, we have focused on the new representational capabilities that handheld computers can bring to the classroom. In conjunction with the first guideline, above, this work led us to realize that handhelds could broaden the range of assessment tasks in which students could express what they know and can do. Thus, we sought actively to exploit the unique representational capabilities of handhelds compared with paper, (e.g., to allow drawing with a stylus, to allow communication and aggregation of student input, to provide interactive feedback) so as to create new ways to capture what students know and can do. Because teachers' existing assessment practices were not supporting their instructional goals, we avoided implementing traditional assessments using handheld technology.

Guideline #3: Handheld assessments should emphasize significant new roles for students.

Our conversations with our Gullah informant about the need to empower students, as well as our observation that many Beaufort teachers were not very comfortable with using technology, led us to another guideline involving significant new roles for students. Everything we had learned suggested that students would more rapidly be able to do new things with handhelds than without them, while teachers would resist converting their existing practices over to a technology-based form. A competing vision, which we came to deemphasize, was that handhelds should bring teachers into more frequent use of the district's technological infrastructure. Our interviews suggested that teachers were likely to resist this vision, and our research context suggested that it would be unlikely to result in better alignment of assessments with inquiry-oriented instructional strategies. An emphasis on students, on the other hand, could be adoptable within Beaufort's cultural and teaching context and could lead to greater instructional effectiveness.

In the design conference, we implemented this guideline by encouraging teachers to focus specifically on self-assessment and to resist linking assessment data with the local accountability system. All students would have the opportunity to use the software designed, and teams were encouraged to imagine how students would be able to reflect on their own learning through use of the tools and how the tools might provide ways to promote better student engagement in the learning process. In addition, despite district officials' interest in doing so, we resisted the idea of integrating classroom-based data with the district's own assessment system, which is used for accountability purposes. We felt that this integration would make it less likely that students would be attracted to the handheld computers, because they would recognize them as just another form of testing.

Guideline #4: Design simple, focused additions to teachers' existing inquiry-oriented practice.

We found that the teachers we observed in transition to inquiry-oriented science teaching were already burdened with many demands to manage their attention and the attention of their students in the classroom. Handheld assessments could help only if they minimized the need to attend to new management tasks and emphasized a focus of attention on central science concepts. To put it simply, it is already too common that students are "lost" in science inquiry tasks; we could not risk complex technology interfaces that might provide a new place to become lost. Further, teachers could not afford to divert their attention from students and science concepts to managing technology. Therefore, we decided it was a necessity to make simple, focused designs.

Guideline #5: Design social activities in the classroom to help students and teachers use representations to assess what they know and can do.

Handheld representations can provide students with new media and mechanisms for expressing what they know; these representations can become useful as assessment data, however, only when they inform further reflection and classroom discussion. We discovered early in the design process that along with specific handheld functionalities, design teams also needed to develop social activities in the classroom to support specific assessment functions in the classroom. These include providing instructions to students at the beginning of activities that set guidelines for quality in student work, rubrics to use in judging the quality of student representations, and formats for group discussion of student work. Although such activities do not require the use of handheld computers for teachers to orchestrate them, they are greatly aided by the kinds of representations handhelds can produce. Furthermore, the artifacts produced on handhelds can be revised and revisited, a critical element in supporting students learning from assessment activities (see National Research Council, 1999).

On reflection, we have found that Zuboff's contrast between "informating" and "automating" IT implementations is a concise way to summarize the impact of these four guidelines. Automation is clearly not called for by our requirements analysis because of the mismatch between what teachers do now and those assessment practices that could improve the effectiveness of their instruction. On the other hand, we could see these guidelines as pointing to the need for handheld assessments to "informate" science inquiry instruction—that is, to provide new representational means that enable students to better develop and demonstrate their competence, to help make teachers more aware of what students know and can do, and to increase the focus throughout the classroom on important science concepts and what it means to demonstrate understanding of them.

Survey of existing handheld assessment products

Before designing new handheld assessment tools, as part of our requirements analysis we also surveyed existing products to see if any might fit. We found mostly products that automate existing assessment practices. We do not mean to diminish the benefits of these products for some instructional goals (e.g., making teacher grading and recordkeeping more efficient), but our analysis suggests that they are unlikely to help teachers make science investigations a more effective activity for science learning.

There are a number of handheld technologies that focus on automation of assessment. Sunburst Technology's Learner Profile to Go is one example (http://www.sunburst.com). It automates the transfer of information from paper-recorded observations of student behavior to teacher grade book by using the handheld as a collection device. Scantron (http://www.scantron.com) and Kaplan (http://www.kaplan.com) have developed software for handhelds that allows 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 by using a program on their desktop computers. Wireless Generation's handheld assessment software simplifies the data capture and management process for elementary-level teachers who maintain running records of students' progress in reading (http://www.wirelessgeneration.com). Its 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.

Standards development processes (some sponsored by IEEE) have also focused on automating traditional assessment. For example, the IMS Question and Test Interoperability (QTI) specification provides a means to express traditional assessments in platform-independent XML format (http://www.imsglobal.org/question/index.cfm). Likewise, the School Infrastructure Framework (SIF) emphasizes information exchange among legacy student records but does not offer any specific supports for improving science inquiry instruction (http://www.sifinfo.org/). Student profile projects like SIF tend to capture information that is of more use to administrators than to teachers who are attempting to improve instruction. In the long run, convergence will be necessary between products that automate and products that informate; in the short run, however, we believe teachers will realize little improvement in science instruction from starting with the automating perspective. Some parents and teachers may be resistant to adopting handhelds to enable districts to achieve automation, since they may see this more as consolidating central bureaucratic power, rather than empowering students and improving instruction.

GoKnow's Palm Archive and Application Manager (PAAM) is an example of a product that fills a gap between automating and informating perspectives by managing teacher distribution of documents to students and collection of work from them (http://goknow.com/Products). Intrinsically, PAAM is a management tool that automates classroom processes; it is not a source of new tasks that enable students to better express what they know and can do. But because PAAM has an extensible architecture, it can enable management of the new kinds of informating assessments in ways that other products cannot. Further, collection and reflection on student work is critical to our vision, and PAAM makes those easier.

We were not able to find many handheld assessment tools that exemplify our informating perspective. The best example we were able to find was Classtalk, a student response system (Abrahamson, 2000). Classtalk informates classrooms by rapidly assembling a public representation of how the classroom group is thinking about scientific concepts. For example, in the Peer Instruction pedagogy (Mazur, 1997), Classtalk is used to rapidly poll students on questions that are calibrated to bring potential misconceptions to the surface. The poll results are instantly available to the teacher and students in a representation that reveals patterns of difference among students in the classroom. Teachers use this information to take the pulse of the group's level of understanding during class, to direct students to work collaboratively with peers to make sense of particular conceptual difficulties, and to adapt the pace and content of their instruction. We did not, however, adopt Classtalk in Beaufort for two reasons. First, it is better suited to improving lecture-oriented science instruction than inquiry-oriented science instruction. Second, it requires a classroom network, and the Palm computers we were using support only peer-to-peer beaming.

A handheld assessment that informates: The case of Sketchy

Project WHIRL set out to create designs to fit the guidelines and informating perspective, albeit by reusing existing software whenever possible. We worked in close partnership with a group of seven teachers to design or adapt software for handhelds to identify designs that were important to their classrooms, supportive of learning goals, useful and usable in an "everyday" sort of way, and feasible to create and support (Patton, 2002).

One team of teachers worked on adapting an existing piece of software, Sketchy, which had been developed by researchers at the Center for Highly Interactive Computing in Education (hi-ce) at the University of Michigan. Sketchy was selected in our initial workshop when we introduced it to show teachers what handhelds could do. Several teachers immediately began to sketch animated pictures of scientific processes, such as the cycle of rainfall or the growth of a plant. Other teachers recognized the potential of handheld animations to enable them to better see what students know about "sequencing" the steps in a scientific process; sequencing is important in South Carolina's standards for science instruction, and teachers felt it was an area in which handheld assessments could readily give them new, useful information about what their students know and can do.

Users of Sketchy can create, edit, and share color animations of scientific processes by organizing a set of individual drawings (frames) they make directly on the handheld screen with a stylus. Users can choose from a number of drawing tools, similar to a palette of a drawing or paint program. The drawing tools include a color palette, a text writer, and various shapes and stamps. Users can play animations backwards and forwards and at different speeds. They can beam an individual frame or an entire animation to another handheld device with Sketchy. Sketchy makes unique use of the handheld representational medium; drawing sketches is much easier with a stylus than with a mouse, and the computer makes it easy to produce animations of processes that occur over time—a big advantage over paper. There are software tools that we could have selected that have similar functionalities; however, Sketchy was selected because teachers were drawn to it as a starting point for the design process and our project goal was to follow teachers' initial ideas for improving assessment at the outset.

Although Sketchy was originally designed as a tool for instruction, our teachers adapted it for assessment purposes. Using Sketchy as an assessment tool requires that its use be embedded in activities that give students a new medium for reflecting on and expressing what they know. In constructing animations, students are much more involved in the process of self-assessment. They must decide what is important and what is peripheral to the process they are animating and use their decisions to indicate what stays constant (by choosing a background) and what changes (by choosing foregrounds). They need to decide on conventions for representing scale and units of time, as well as the level at which to animate a process (microscopic, observable, etc.). See Figure 1 for an example of a student's sketch.

[-------------------INSERT FIGURE 1 ABOUT HERE-------------------]

Classroom discussions about student-created Sketchy animations are another important occasion for assessment. Teachers can ask students to use a visualization to explain a process they have studied in class; such an activity is an important preparation for scientific inquiry, since visualization is a key aspect of the practice of science inquiry (Latour & Woolgar, 1986). As part of these discussions, teachers can employ rubrics for assessing the adequacy of student animations and the quality of their explanations.

Participating teachers have found Sketchy easy to use. They have also found it fit well with their goals for instruction and assessment. Teachers from all the design teams found it to be the most suitable piece of software to use to introduce students to handhelds; they reported that their students, in turn, really enjoyed using Sketchy in class. It required little setup time, and the most guidance teachers found they needed to provide for using the software was a large picture of the buttons on the handheld computer displayed at the front of the class and a short description of the tools in the palette. One teacher who was part of the Sketchy design team has found having students create and label animations valuable in helping distinguish more quickly among students' degrees of understanding of the difference between a producer and a consumer in a food web. "Because I had them label pictures, I didn't have to worry. Kids did have to explain, 'You're telling me the snake is going to eat the eagle,' because kids were drawing arrows in the wrong direction." Another teacher has found it a useful tool for her to help her school enact its "arts-infused" focus; she used it early in the project to give her students more practice in learning the difference between potential and kinetic energy by having her students draw different animations to illustrate their understanding of the concept. A third teacher reported benefits to using Sketchy to teach her students about the phases of the moon. She observed that her students looked at their textbook's diagram of the phases in a new way—to understand it—in order to create their animations. Her observation is supported by her students' test results; she noted that her students who used Sketchy to animate the moon's phases performed much better on the end-of-chapter test than students in years past.

For these teachers, Sketchy helped to broaden the repertoire of assessment strategies they were using. Creating animations involved students in more direct, active encounters with subject matter than textbook diagrams or descriptions. They succeeded in engaging students in commonly assigned tasks—such as reading the textbook—in new ways that resulted in better retention of knowledge. And they expanded the range of means of self-expression for the students involved. Students who had before been judged "low-performing" on the basis of traditional assessments found new ways to engage in learning and demonstrate their competence to their teachers, who in turn recognized what they knew and could do.

Conclusions

Improving assessment is an important opportunity for the application of handhelds to education. In this paper, we have focused on analyzing the requirements for realizing this opportunity with respect to inquiry-oriented science instruction, using data from a fairly normal, diverse school district. These requirements have led us to emphasize a need for handheld-based tools that "informate" science instruction by

- Being oriented to the needs of teachers in transition to inquiry-oriented pedagogy;
- emphasizing expansion of the range of assessment tasks through a new representational medium and communication infrastructure;
- creating new roles for students in expressing what they know and can do; and
- focusing both students' and teachers' attention on scientific concepts.

An evaluation of Sketchy and two other tools designed within an informating perspective is currently under way. In future work, we would encourage the community of designers and researchers in this area to attend to a balance of informating and automating perspectives as they work to add assessments to their designs and advance the effectiveness of handhelds in improving learning.

Acknowledgments

This work has been supported by National Science Foundation grant #0126197. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. We are grateful to the Beaufort teachers, school district staff, and community members who have worked closely with us. We thank Elliot Soloway and the University of Michigan for making Sketchy available to our project, and Esther Verreau for her many improvements to it.

References

Abrahamson, A. L. (2000, March). *A brief history of Classtalk*. Paper presented at the T3 International Conference, Dallas, Texas.

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.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213-238.

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 the 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. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Blumenfeld, P., Soloway, E., Marx, R. W., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26(3&4), 369-398.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245-281.

Chen, Y.-S., Kao, T.-C., Sheu, J.-P., & Chiang, C.-Y. (2002, August). *A mobile scaffolding-aid-based bird -watching learning system*. Paper presented at the International Workshop on Wireless and Mobile Technologies in Education, Växjö, Sweden.

Cognition and Technology Group at Vanderbilt (1997). *The Jasper Project: Lessons in curriculum, assessment, and professional development*. Mahwah, NJ: Erlbaum.

Cole, K. A., Coffee, J., & Goldman, S. V. (1999). Using assessments to improve equity in mathematics. *Educational Leadership*, 56(6), 56-58.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.

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

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathwaway (Ed.), *New directions for testing and measurement: Testing in the schools* (Vol. 19). San Francisco: Jossey Bass.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208.

Kaput, J., & Hegedus, S. (2002). *Exploiting classroom connectivity by aggregating student constructions to create new learning opportunities*. Paper presented at the 26th Conference of the International Group for the Psychology of Mathematics Education, Norwich, UK.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.

Mazur, E. (1997). *Peer Instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

Means, B., & Haertel, G. D. (2002). Technology supports for assessing science inquiry. In Board on Testing and Assessment (Ed.), *Technology and assessment: Thinking ahead*. Proceedings from a workshop (pp. 12-25). Washington, DC: National Academy Press.

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. 1. Methods of Evaluating Educational Technology*. Greenwich, CT: Information Age Press.

Millen, D. R. (2000). *Rapid ethnography: Time deepening strategies for HCI field research.* Unpublished manuscript.

National Research Council. (1999). *How people learn: Brain, mind, experience, and school.* 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). *Knowing what students know*. Washington, DC: National Academy Press.

National Research Council. (1996). *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.

Patton, C. (2002). *Project WHIRL design rubric*. Menlo Park, CA: SRI International.

Petrosino, A. J. (1998). *At-risk children's use of reflection and revision in hands-on experimental activities*. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.

Rudd, T. J., & Gunstone, R. F. (1993, April). *Developing self-assessment skills in grade 3 science and technology: The importance of longitudinal studies of learning*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta, GA.

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.

Solomon, G., & Perkins, D. N. (1998). Individual and social aspects of learning. *Review of Research in Education*, 23, 1-23.

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.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practice of classroom assessment*. Albany, NY: State University of New York Press.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Stroup, W. M. (2002, September). *Instantiating Seeing Mathematics Structuring the Social Sphere (MS3): Updating generative teaching and learning for networked mathematics and science classrooms*. Paper presented at the International Conference of the Learning Sciences, Seattle, WA.

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.

Yarnall, L., Penuel, W. R., Ravitz, J., Murray, G., Means, B., & Broom, M. (2003). Portable assessment authoring: Using handheld technology to assess collaborative inquiry. *Education, Communication, Information*, 3(1), 7-55.

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.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York: Basic Books.

Figure 1. A student's Sketchy animation of the food cycle near a pond.