Evaluation of a Handheld Data Collection Interface for Science Learning

Cynthia Sims Parr,[1,3] **Tricia Jones,**[1] **and Nancy Butler Songer**[2]

Despite a rise in the use of handheld computers in classrooms, meaningful learning with personal digital assistant (PDA) technology remains poorly studied. We report results from an evaluation of customized handheld data collection software, the BioKIDS Sequence, which was used during an 8-week biodiversity curriculum unit by 5th and 6th grade students in southeastern Michigan. We provide new information on design decisions and usability of our customized software; discuss learner use and preferences; and report rates of data accuracy. Most students were able to enter simple animal observations using an icon-driven, largely linear interface. The interface helped expand the types of data students were aware of, and allowed them to view and review their entries and flag them as uncertain, factors important in supporting accurate data collection. Other BioKIDS research indicates students were subsequently able to see simple patterns in their data to guide the formulation of future hypotheses, questions, or conclusions. The BioKIDS Sequence and PDA technology therefore represent a meaningful use of technology to support scientific reasoning.

KEY WORDS: interface; portable; inquiry learning; biodiversity; accuracy; usability.

INTRODUCTION

As use of handheld computers in the form of personal digital assistants (PDAs) increases, classroom-based meaningful learning with PDA technology remains poorly studied (Songer, in press). Recent research documents that teachers using handhelds for inquiry-based instruction see potential benefits (Crawford *et al.*, 2002), but very little research exists so far on the effectiveness of PDAs for specific educational practices. Researchers and software designers have only begun to address how to custom-design PDA-rich learning environments for particular learning goals, purposes, or audiences.

One emerging use of PDAs for meaningful classroom-based learning is for data collection by students. Handheld data collection is a faster and more accurate method for transferring data than paper-based methods (Hecht, 1997; Spain *et al.*, 2001). Moreover, data entry via a pen-based PDA involves less technology anxiety than other forms of computer-aided data collection (Tseng *et al.*, 1998). Professionals in a variety of settings recognize these advantages and encourage use of PDAs for data collection in a range of contexts (e.g., Weber and Roberts, 2000; Greene, 2001; Hampshire, 2001).

In the field of science education, current national standards emphasize the development of complex reasoning skills in concert with the development of conceptual understandings of scientific phenomena. Scientific reasoning skills include building on children's natural curiosity towards collecting accurate scientific data, analyzing data, asking questions, and building explanations of scientific phenomena (National Research Council, 2000). In particular, standards recommend the use of "appropriate tools and techniques to gather, analyze, and interpret data" (National Research Council, 1996). Previous research demonstrates that when computers are

[1]Museum of Zoology, The University of Michigan, Ann Arbor, Michigan.
[2]School of Education, The University of Michigan, Ann Arbor, Michigan.
[3]To whom correspondence should be addressed at A. V. Williams Building, University of Maryland, College Park, Maryland 20742; e-mail: csparr@umd.edu

used as one of these appropriate tools to gather analyze, and interpret data, students can prioritize data analysis and pattern discovery as opposed to the logistics of the data collection process itself (Brasell, 1987; Thornton and Sokoloff, 1990).

Handheld data collection tools may also provide ways for students to contribute meaningfully to citizen science projects such as stream, bird, and amphibian monitoring. To develop tools to support accuracy in student data collection, we must first measure baseline error rates and then test software designs that might improve students' ability to collect accurate data. Few estimates of childrens' datataking or data entry accuracy are available. One GLOBE project found 67% accuracy in coding of land cover classes (Becker *et al.*, 1998); other GLOBE evaluations focus only on implementation consistency and not on other factors influencing data quality (Howland and Becker, 2002; Penuel, 2002).

In a previous paper we described the design of our handheld data collection software (Parr *et al.*, 2002). In this paper we report results from an evaluation of this software which was used during an 8-week curriculum unit by 5th and 6th grade students. We provide new information on design decisions and usability of our customized software, learner use and preferences, and data accuracy. The paper concludes with a discussion of the role of our tool in supporting the development of scientific reasoning.

RESEARCH METHODS AND RESULTS

Curriculum Design

The primary goal of the BioKIDS: Kids' Inquiry of Diverse Species project (www.biokids.umich.edu) is to support fifth through eighth grade students in failing Detroit Public Schools to demonstrate complex reasoning in science and technological fluency as a result of their interaction with a challenging, multi-year, coordinated science program. Using science units designed to promote complex reasoning, students utilize current technologies such as handheld computers to collect scientific data in a manner similar to professional scientists. Once data are collected, students analyze data and generate scientific claims, explanations, and hypotheses based on their data (Songer *et al.*, 2003). To date, several cohorts of Detroit students have demonstrated significant gains in complex reason-

ing associated with biodiversity data (e.g. Songer, 2003).

Interface Design

Essential to the success of handheld data collection is interface design adapted for target audiences and purposes. We customized existing Windows/ Palm OS software for this project. CyberTracker (www.cybertracker.co.za) was originally designed to assist illiterate but highly skilled South African trackers in monitoring wildlife in nature preserves. This software's icon-driven interface seemed well suited to young students, even those with low reading skills or limited English fluency. In addition, we speculated that students might be motivated to learn science through an authentic computer tool used by professional African animal trackers.

The CyberTracker application allows relatively easy customization of data collection screen order and appearance. Our customization, the "BioKIDS Sequence," (Fig. 1) was guided by several design principles which are assessed in this paper. (1) Icons are preferred because selecting from menus and icons can be more efficient than text entry (Preece *et al.*, 1994). (2) Navigation of screens should be largely linear, not branching, and screens should be tightly coupled with curricular goals (Luchini *et al.*, 2002; Parr *et al.*, 2002). We departed from specific curricular needs only in including several screens related to observation type and data quality. (3) Screens should represent taxonomic common sense (Parr *et al.*, 2002). In other words, the BioKIDS Sequence adhered to accepted animal classification at the top level to reinforce curricular content. To improve usability by students, however, at lower levels we used quick lists of commonly seen animals.

Study Design

Our study had three parts, each featuring different but complementary data. In Part 1, we collected and analyzed CyberTracker data logs from eight classes of primarily urban 6th graders ($n = 163$ students). In Part 2, we administered student opinion surveys ($n = 45$) and conducted one-on-one interviews ($n = 6$) at the completion of the 8-week biodiversity unit. In Part 3, we collected and analyzed Cyber-Tracker data logs from a technology practicum exam administered to students both before ($n = 346$) and
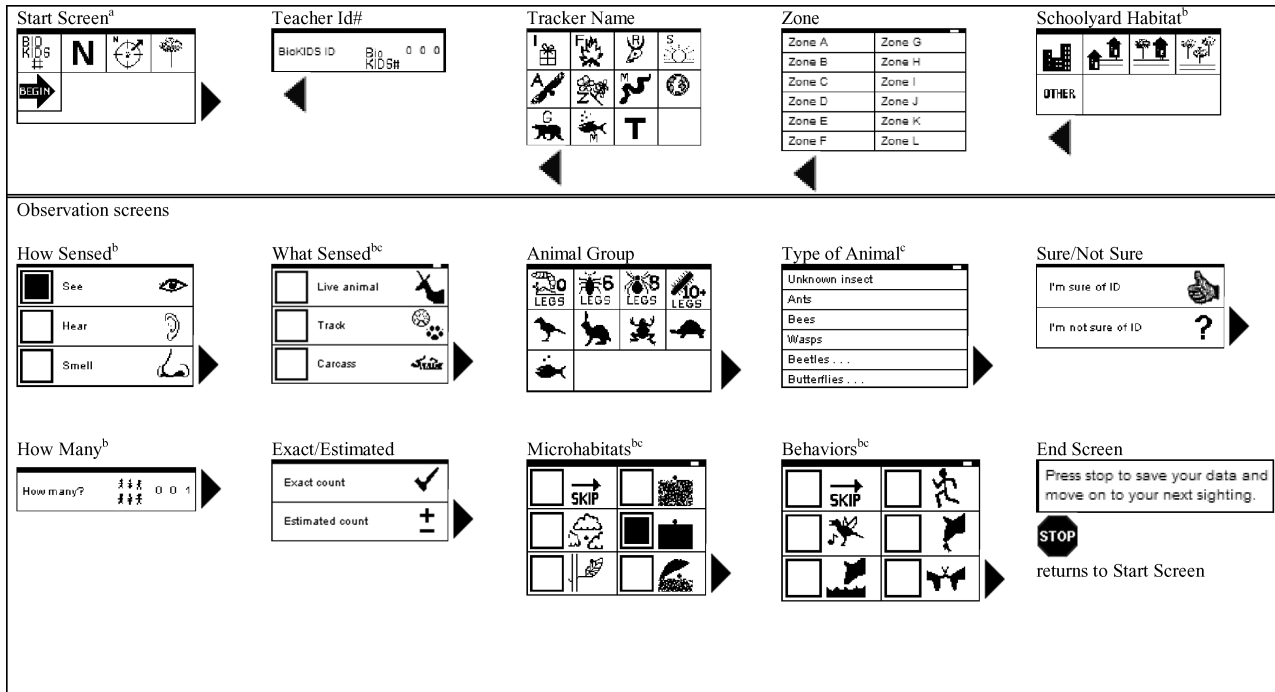
Fig. 1. The BioKIDS Sequence. (a) *Start* screen is a branch point of four "state screens" whose state is retained, unless changed, throughout a session. The "Begin" arrow on *Start* screen is used to get to *Observation* screens, which are always presented in the same order. After each screen, users press a "next" arrow to proceed; backing up to reenter data is possible until they select "Stop." A history bar (not shown) shows student choices for four screens at a time. (b) Screen can be bypassed without a choice. (c) Screen has more than one page of choices.

after ($n = 198$) the next enactment of the biodiversity unit.

PART 1: CURRICULUM CYBERTRACKER LOGS

Methods

Classroom-Based Practice Activity

To gain familiarity with the interface, students first entered data using the BioKIDS Sequence for four unlabelled photos of animals and, as time allowed, for pictures in classroom books. Students used specially created ID guides to identify animal species. Data consisted of four or more entries per pair of students, totaling 202 entries from at least 43 student pairs.

Fieldwork Activity

After the Practice activity, students used the BioKIDS Sequence to collect 278 field-based obser-

vations. Students worked in teams of four, with one student at a time recording data.

Analysis

One way to assess usability is to consider whether any screens were not used. Some screens could be bypassed (see footnoted screens in Fig. 1) by tapping on the "next" arrow without making a choice. On some screens students could also choose "Skip." The *How Many* screen could also be bypassed by leaving the number at "000." In our analysis, we coded an entry as incomplete if any screen was bypassed or skipped.

In the Practice activity, we coded a screen as incorrect if students entered any data that did not match information in the practice photos. An overall entry was incorrect if data from at least one screen was incorrect. In the Fieldwork activity, a screen was coded as inaccurate if students entered data that represented either implausible or unlikely numbers, species, microhabitat and/or behavior as determined by our field biologist. We used this standard as we

had no independent confirmation of actual student sightings. An example of an implausible entry might be a mourning dove observed "moving and drinking" while in the microhabitat "in or on dirt." Given the schoolyard location, this entry would have an implausible behavior (drinking) but the animal and the microhabitat are considered plausible. We considered an entry "uncertain" if a student indicated "I'm not sure of ID" and/or indicated "Estimated count" after entering *How Many*. We considered an identification "vague" if they chose "other" or "unknown" on any *Type of Animal* screen.

Results

Usability

Students fully completed 46% of Practice activity entries and 39% of Fieldwork activity entries. Of six screens that could be bypassed or skipped, two screens *What Sensed* and *How Sensed* were bypassed less than 3% of the time. Students skipped the *Schoolyard Habitat* screen 21% of the time during fieldwork, perhaps because they had been instructed to skip this screen in the practice activity as instruction had not yet been provided on habitat classifications. *Microhabitat* was bypassed (41%) or skipped (11%) in half the fieldwork entries. *Behavior*, which uses a similar interface style as *Microhabitat*, was bypassed (8%) or skipped (19%) much less often. Thus, the bypassing of *Microhabitat* is probably not due to the usability of the interface style but due either to difficulty in determining the habitat, lack of interest in those data, or difficulty understanding the icons.

Accuracy

About 22.28% of Practice entries were incorrect or implausible. The number of clearly implausible entries drops significantly ($p < 0.001$) to 8.63% in Fieldwork. Even when "000" *How Many* entries are excluded, most of the 91 errors were on the *How Many* screen (Table I). Errors in *Teacher ID*, which uses the same input method as *How Many*, had a much lower error rate. Note however, that students only had to enter the *Teacher ID* value one time for all entries during a session.

Student data reflects a somewhat surprising ability to notice a range of types of animal sightings beyond live animal sightings, such as animal tracks or

Table I. Kinds of Errors Made (Implausible or Incorrect) in Practice and Fieldwork Activities

Kind of error	Practice (%)	Fieldwork (%)
Animal Group or *Type of Animal*	19	24
How Many	46	41
Behavior	16	29
Teacher ID	12	3
Microhabitat	7	3

signs. Twenty-five percent of Fieldwork entries included carcasses, tracks, signs, smell or sound as a part of observations.

Many students indicated uncertainty both in Practice (32%) and Fieldwork (23%), and this decrease of uncertainty with experience is statistically significant ($p < 0.05$). Ten percent of Practice and 9% of Fieldwork entries were both implausible and uncertain. Entries with uncertainty were not significantly more likely to be implausible than those marked certain. However, if the data are pooled across Practice and Fieldwork activities, students were particularly likely to flag their data as uncertain if they had multiple mistakes ($p < 0.05$). Removing from the analysis all data that students flagged as uncertain did not change the overall percentage of implausible entries. Multiple fieldwork "errors" may actually represent bored students making up entire observations when no animals were present (A. Huber, pers. comm.).

The percentage of vague identifications was higher in Fieldwork as compared to Practice. Almost 40% of all observations had identifications that could have been entered into the software more specifically. Vague identifications were significantly unlikely to be coded implausible in both Practice and Fieldwork ($p < 0.05$).

PART 2: SURVEY AND INTERVIEWS

Methods

We administered a questionnaire with nine Likert-format questions and three open-ended questions to 45 students in order to evaluate student beliefs about the BioKIDS Sequence. We analyzed the Likert questions using simple statistical methods. We coded open-ended questions using a researcher-generated coding rubric. To analyze student screen preferences, we clustered all screens by theme and by interface style and compared student responses to the frequency expected if each screen in a sequence were equally preferred.

Table II. Survey Results ($n = 45$ Students)

Statement	Responses (%)		
	Agree	Not sure	Disagree
I had a hard time learning how to use CyberTracker	17.8	20.0	62.2
The icons helped me enter data faster	66.7	20.0	13.3
I did not really notice the history bar at the top	46.7	37.8	15.6
Using CyberTracker was fun	80.0	13.3	6.7
Most of my CyberTracker entries were correct	60.0	33.3	6.7
Sometimes I didn't know what to tap on	44.4	20.0	35.6
I paid more attention to animals because of CyberTracker	40.0	40.0	20.0

After the survey, we asked six students to expand on their written survey responses during one-on-one interviews consisting of 15 questions. These 10–15 min interviews included a think-aloud portion where the student created an entry. We videotaped, transcribed, and coded interviews following traditional qualitative analysis procedures (Miles and Huberman, 1994).

Results

Usability

Survey results indicate that the majority of students found the BioKIDS Sequence easy to learn and fun to use (Table II). These results support our hypothesis that an icon-driven interface would be helpful for young students, as 66% agreed that the icons helped them enter data faster. Table II results also indicate that students sometimes had difficulty knowing what to tap on.

Table III presents results to the survey question about the most confusing aspects of the BioKIDS Sequence. Some students mentioned trouble understanding the icons, or requested more text with the icons. Some students mentioned navigation as an area

Table III. Most Confusing Aspects of the BioKIDS Sequence ($n = 33$ Students)

Category of response	Responses (%)
General (learning or knowing what to do)	27.3
Icons/text	24.2
Navigation	18.2
All other specific CyberTracker features	15.2
PDA technology issues	15.2

of confusion. In particular, students wrote they had difficulty knowing what to tap at times, and "how to keep going forward." One student mentioned that he found the arrows at the bottom of each screen confusing, while another said that she wasn't sure what to do when she didn't understand the icons.

Animal and tracker name icons were particularly compelling for students. When asked "What is your favorite icon?" students with a preference overwhelmingly named an icon with an animal pictured (56% of 32 students). Seven students responded with a tracker icon. One student we interviewed said he liked the frog icon because it was his favorite animal. Another liked the singing bird (a *Behavior* icon) because of the music notes, while a third liked the bird icon on the animal group page because the icon was "real obvious, nice, and bold." Four students remembered the icon well enough to spontaneously draw it when asked about favorites on the survey. Similarly, when asked, "What was your favorite screen?" students preferred screens consisting of icons, and those in the thematic groups Animals and Extras (Table IV). Three students mentioned "picking our tracker name" for favorite screen. These results, taken together with the popularity of tracker icons, suggest that selecting a tracker name—even though it happens only once per session—is a compelling aspect of the interface for some students.

Students provided suggestions for how to improve the BioKIDS Sequence. In general, a majority of students believed more words coupled with icons would enhance the understandability of the icons for a wide audience of users (Table V). Though this appears contrary to their preference for icons, students said in interviews that they did not want words to replace the icons, but just more labels and directions on the icon-rich screens. CyberTracker software supports a "tap and hold" feature that allows students to get text help for any icon. Once they were aware of this feature in the BioKIDS Sequence, most but not all students became adept at interpreting the icons.

Accuracy and Proficiency

Most students were fairly confident their BioKIDS Sequence entries were accurate though a third were not sure (Table II). Independent sample t-tests showed that students who had a hard time learning the BioKIDS Sequence (answered "agree" or "not sure" to the first statement in Table II) were significantly less confident about their entries ($p < 0.05$)

Table IV. Analysis of Screens Designated as "Favorite" by Students ($n = 27$ Students)

	% Expected[a]	% Favorite screen
Thematic group		
Tracker (*Tracker Name*)	6.7	11.1
Other start (*Start Screen, Teacher ID, Zone, Schoolyard Habitat, End*)	33.3	29.6
Animal (*Animal Group, Type of Animal*)	13.3	22.2
Sensing (*How Sensed, What Observed*)	13.3	7.4
Numbers (*Sure/unsure, How Many, Exact/estimate*)	20.0	7.4
Extra (*Microhabitat, Behavior*)	13.3	22.2
Interface style		
Icons (*Tracker Name, Animal Group, Behaviors, Microhabitat*)	40.0	74.1
Icons and text (*How Sensed, What Sensed, Sure/unsure, Exact/estimate*)	40.0	14.8
Text (*Type of Animal*)	20.0	11.1

[a] % expected is the actual distribution in the BioKIDS Sequence.

and were less confident about how to use the sequence ($p < 0.01$) compared to those who found the BioKIDS Sequence easy to learn.

In interviews, students showed a variety of methods for dealing with uncertainty, including vague identification ("unknown insect") and skipping. One student used two different strategies in the same entry: he changed his microhabitat choice to "other" while stating he was not sure, but then went on to skip the behavior screen because "I don't know what they're doing so I'll put skip." Two students asked the interviewer for help identifying the microhabitat. When asked what they would do in the field when unsure, students said they would use "skip," "other," or "unknown," consult a field guide, or ask a teammate or teacher for help. All but one student demonstrated how to step back through an entry to correct errors.

Student Views on Learning With the BioKIDS Sequence

Surveys and interviews also assessed student views on what they learned using the BioKIDS Sequence. Forty percent of students said they paid more attention to animals because of CyberTracker (Table II). When asked "Something I learned by using CyberTracker was . . . ," 46% refer to technology (e.g. "how to use CyberTracker" or "I learned more about using palm pilots"), while 54% mention specific

Table V. Opinion of BioKIDS Sequence Interface Styles ($n = 45$ Students)[a]

Response includes "Needs more pictures"	24.3%
Response includes "Needs more words"	53.2%
Response includes "Okay as is"	39.2%

[a] Responses add up to more than 100% because students could choose more than one statement.

science content. Of the content responses, almost all have to do with animal observation or identification (e.g. "I practiced observing things," "I paid a lot more attention to animal behavior," and "there are more animals than I thought there were").

In interviews, all six students mentioned specific animal content they learned while using the BioKIDS Sequence. One student declared he hadn't noticed how many animals or species were in the schoolyard prior to this experience—he was made aware by entering them with CyberTracker and by seeing class totals afterwards. Three of the students mentioned the summative animal lists and how the lists made them aware of different species in the schoolyard. Three students mentioned that the BioKIDS Sequence helps them be a better observer because it focused their attention on what to observe. For example, "It has steps, so you know to say where it lives and what kind of animal it is. So you remember to describe how it looks and where it is. Because then it helps you remember."

PART 3: PRACTICUM CYBERTRACKER LOG ANALYSIS

Methods

In order to measure usability and accuracy rates of the BioKIDS Sequence under controlled conditions, we assessed student performance on a technology practicum exam. The BioKIDS Sequence was simplified and modified for the practicum exam to preserve partial entries (Fig. 2).

Pretest Practicum

We administered the practicum before the BioKIDS curricular unit. Students used the handhelds

Start Screen (no data entered) → Animal Group → Type of Animal (up to two screens) → How Many → Microhabitat → End Screen

Fig. 2. The BioKIDS practicum sequence. (a) See Fig. 1 for screenshots. In this version, each screen had a stop icon to allow partial entries to be saved.

to create an entry for a labeled animal photograph (one jumping spider on something hard). They were given no instructions on PDA use or CyberTracker software. Each student had about 5 min to complete the entry.

Posttest Practicum

After completing the 8-week BioKIDS curricular unit, the same practicum exam was administered. By the posttest, students were familiar with the PDAs and CyberTracker through participation in BioKIDS.

Analysis

Data consist of one entry from each student who took the pretest ($n = 346$ entries) and posttest ($n = 198$) practicum exam. All entries were evaluated for completeness and accuracy. If a student stepped through, or used, all screens including microhabitat the entry was considered complete whether or not accurate data had been entered. Only the number screen could be skipped; for all other screens a failure to step past a screen meant that no further screens could be attempted. Screens that were skipped or not used were not included in calculations of accuracy. Data were scored as errors if they didn't match the label on the photograph exactly. For example, since we know the *How Many* value should be "001," a value of "100" would be an error. The photograph labels indicate a single microhabitat, so multiple microhabitats were considered an error even if the students included the single correct microhabitat. An entry was scored inaccurate if at least one screen had an error.

Results

Usability

Overall, results suggest that the BioKIDS Sequence was easy to use, especially after a little expe-

rience. Even without any instruction, almost a third (27%) of the students completed the entire entry on their first exposure in the pretest practicum. Completion rate increased significantly to 57% by the posttest ($p < 0.001$). However, the number of students who did not use the sequence at all remained the same (14% pretest, 13% posttest).

To determine which screens were immediately usable and which required training and experience, we compared screen usage rates during the pre- and posttest practicum (Table VI). While a high percentage of students used the *Animal Group* screen, the *Type of Animal* quicklist screens posed some difficulty for students in the pretest. Note that a wrong choice on the *Animal Group* screen would mean the right choice was not apparent on the next *Type of Animal* screen. Many students were unable to use the *How Many* screen, both before and after the curriculum. Quite a few students do not use the *Microhabitat* screen in the pretest but in the posttest almost all who used *How Many* also used *Microhabitat*.

Accuracy

Overall accuracy in the practicum improved significantly with curricular experience: 43% of attempted entries were entirely correct up to the point of completion in the posttest practicum compared to 29% correct in the pretest practicum ($p < 0.005$). The

Table VI. Percentage of Students Using Screens in Technology Practicums[a]

Screens used	Pretest ($n = 314^a$)	Posttest ($n = 172^a$)
Animal Group	100%	100%
Type of Animal 1	85%	94%
Type of Animal 2	83%	93%
How Many	57%	76%
Microhabitat	44%	74%

[a]Sample includes only students who initiated the BioKIDS Sequence. Although the final screen was *Microhabitat*, not all students using this screen completed their entry because some had skipped the *How Many* screen.

Table VII. Kinds of Errors in Technology Practicums

Kind of error	Percentage of errors made	
	Pretest	Posttest
Animal Group or *Type of Animal*	31	36
How Many	41	27
Microhabitat	29	37

improvement was due to a relative decrease in errors in the *How Many* screen (Table VII, overall χ^2 $p <$ 0.05). About 56% of the *Animal* entry errors involved a wrong selection from the *Animal Group* screen (e.g. insect instead of arachnid), and the rest involved errors within quicklists (e.g. mite instead of spider, or funnel-web spider instead of jumping spider, even though jumping spider was written on the photo).

About a third of the entries had multiple errors (Fig. 3). The distribution of numbers of errors changed significantly from pre- to posttest practicum ($p < 0.01$). The percentage of students making two or three errors remained essentially the same; gains in accuracy appear to be the result of fewer students making a single mistake.

Discussion

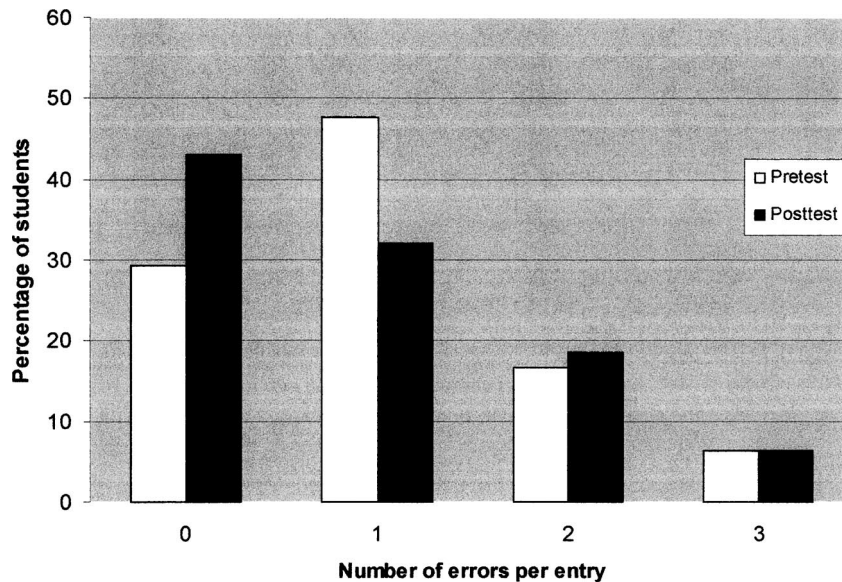### Usability

Our results suggest that the BioKIDS Sequence is quite usable for students in our target audience.

Almost 30% of students could record a sighting the first time they picked up the PDA. This is not entirely surprising, as the original CyberTracker program was already usable for its target audience of professional wildlife trackers. This moderately high success rate for first-time usability suggests that the software is well-suited for data recording even for young students with essentially no field experience.

The practicum improvements underscore how students can learn the software with some experience. Students increase speed (improvements in overall completion) and learn how to use specific screens (improvements in number entry). Some gains might also be a factor of increased content knowledge (e.g. improvement in *Animal Screens* completion). The improvement suggests that our training is adequate, although the difficulties that remain highlight specific problem spots.

The most obvious difficulty is number entry. To enter a number in the *How Many* screen, users must know to tap above each digit independently to increase it and below to decrease it. Some students changed the leftmost digit to a 1, resulting in 100 rather than 001. Although it is hard for many students to figure out this representation without training, the format proved to be largely functional once students learned how to use it. In the future we plan to test an alternative input method that looks and behaves more like a calculator display; this may reduce the error rate substantially if errors are largely a function of the unfamiliar interface.

Fig. 3. Rates of multiple errors per entry in pre- and posttest practicum.

Our three design principles contributed to the usability of the BioKIDS Sequence are as follows.

Use of Icons. Icons are powerful and offer advantages in interface design (Preece *et al*., 1994) but also rely heavily on a shared cultural, age-appropriate context (Soreanu, 1998). These general findings were true of our interface as well, as some icons confused some students. However, overall, students succeeded with the icon-driven interface, and they like the icons. Most students seemed to recognize the value of icon-driven interfaces, yet they asked for more supporting text to clarify icon representations or instructions.

Linear Navigation. We believe that the high completion rates and relative lack of navigation-related complaints suggest the linear navigation is valuable—students are not getting lost, and they consider CyberTracker fun, easy to use, and efficient. Students bypassed and skipped screens easily, although this may have been more a function of uncertainty than efficiency.

Taxonomic Common Sense. Students considered the top-level distinction between animal groups to be clear and useful. They also like the animal lists because they serve as a visible reminder of the previously hidden animal diversity in their schoolyard. Students consider the combination to be efficient for recording live animal sightings. Furthermore, the practicum improvement in *Animal Screens* completion suggests that students are not confused by the presentation.

Accuracy for Contributing to Real Science

Although our goal was not to contribute data to actual ecological studies, we discuss below three ways that the BioKIDS Sequence effectively assists accurate data collection. First, it allows students to skip the step of transferring data from paper to computer, a process known to be error-prone in adults. Second, the BioKIDS Sequence design provides reminders of what students should look for in making accurate observations. Third, the BioKIDS Sequence provides easy ways for students to flag vague or questionable data.

Our study provides some of the first estimates of accuracy in data entry by children. Error rates of trained users in our study ranged from 8.5% implausible entries during fieldwork observations to 56% of entries made in the posttest practicum. Error rates of 1–6% have been measured for adults transcribing written data (reviewed in Blumenstein, 1993). Though our results are not directly comparable to these, as our students transcribed the same observation, rather

than hundreds of observations, we are aware of no other estimates of data entry error among children. The 20% improvement in data entry accuracy from pre- to posttest practicum is encouraging given the limited experience students had with the tool. If the constant percentage of multiple error entries is due to a small number of technophobic or bored students, it may be valuable to identify such children and provide longer training periods or different tools, or to flag their data as problematic.

An accurate entry requires both knowing what data to enter (determination accuracy) and properly entering that data into the device (data entry accuracy). While we did not measure it directly in this study, we expect determination accuracy to be low because our curriculum provides only limited training on animal identification. Future studies should address this. At the same time, our results suggest that students appreciate the supports provided to improve their observation skills. Students often recorded tracks and signs as well as live animals, and interviews indicated they appreciated the lists of animals they were likely to see. Thus, students did use the BioKIDS Sequence to scaffold their observations.

Students also took advantage of screens we provided for indicating uncertainty, essentially flagging their own data as questionable. Students who made multiple mistakes appropriately indicated uncertainty. However, training should emphasize the need to reduce vague and uncertain observations, as these are likely to be discarded from real ecological studies.

CONCLUSION

Our research results suggest that most students were able to learn to use the BioKIDS Sequence as an "appropriate tool and technique to gather, analyze, and interpret data" (National Research Council, 1996). The BioKIDS Sequence and the PDA technology therefore present a case for meaningful use of technology to support scientific reasoning. At the onset, we wished to examine whether or not the technology would play essential roles in minimizing the complexity of accurate data collection, fostering the development of strong observation skills in the field, organizing field data in simple formats, and guiding students to understand key scientific concepts such as animal grouping. With the BioKIDS Sequence, most students were able to enter simple animal data, expand the types of data to be aware of, view and review their entries, and as reported by Songer (2003),

see simple patterns in their data to guide the formulation of future hypotheses, questions, or conclusions.

A primary research goal of the BioKIDS curricular program is to develop strong learning tools including technological tools, and to provide evidence to suggest how the learning tools support the development of complex reasoning in science (Songer *et al.*, 2003). Our results presented here on student usability in the field and on the technology practicum, combined with demonstration of student gains on measures of complex reasoning in biodiversity (e.g. Songer, 2003), begin to provide evidence supporting student use of technology and describing how it can help achieve our desired learning goals. We need to continue to fine-tune our assessment instruments to provide more sensitive measures to better characterize the role of the technology in fostering complex reasoning in science. We believe our results suggest a case for additional research-driven, iteratively-designed learning environments that support accurate data collection in pursuit of standards-driven learning goals. Such powerful uses of technology should help students develop both complex reasoning skills in science and fluency with emerging technological tools.

ACKNOWLEDGMENTS

REFERENCES

Becker, M. L., Congalton, R. G., Budd, R., and Fried, A. (1998). A GLOBE collaboration to develop land cover data collection and analysis protocols. *Journal of Science Education and Technology* 7: 85–96.

Blumenstein, B. (1993). Verifying keyed medical research data. *Statistics in Medicine* 12: 1535–1542.

Brasell, H. (1987). The effect of real-time laboratory graphing on learning graphic representations of distance and velocity. *Journal of Research in Science Teaching* 24: 385–395.

Crawford, V., Vahey, P., Lewis, A., and Toyama, Y. (2002). *Palm Education Pioneers Program March 2002 Evaluation Report*, SRI International.

Greene, P. D. (2001). Handheld computers as tools for writing and managing field data. *Field Methods* 13: 181–197.

Hampshire, V. (2001). Handheld digital equipment for weight composite distress paradigms: New considerations and for rapid documentation and intervention of rodent populations. *Contemporary Topics in Laboratory Animal Science* 40: 11–17.

Hecht, J. B. (1997, July 22–26). Using a PDA for field data collection. Paper presented at the *Annual Meeting of the Mid-Western Educational Research Association*, October 1997, Chicago, IL.

Howland, D., and Becker, M. L. (2002). GLOBE—The science behind launching an international environmental education program. *Journal of Science Education and Technology* 11: 199–210.

Luchini, K., Quintana, C., Krajcik, J., and Soloway, E. (2002, October 23–26). Designing learner-centered scaffolded tools for handheld computers. In Bell, P., Stevens, R., and Satwicz, T. (Eds.), *Keeping Learning Complex: The Proceedings of the Fifth International Conference of Learning Sciences*, University of Washington, Seattle, WA, pp. 268–275.

Miles, M. B., and Huberman, A. M. (1994). *Qualitative Data Analysis*, Sage, London.

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

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

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

Penuel, W. R. (2002). Explaining implementation variation and fidelity in GLOBE: An analysis of student data reporting patterns. Paper presented at the *Seventh Annual GLOBE Conference*, Chicago, IL.

Preece, J., Rogers, Y., Sharp, H., and Benyon, D. (1994). *Human–Computer Interaction*, Addison-Wesley, Boston.

Songer, N. B. (2003). Coordination of the development and evaluation of complex reasoning in science. Paper presented at the *American Educational Research Association (AERA) Annual Meeting*, Chicago, IL.

Songer, N. B. (in press). Curriculum-focused professional development: Addressing the barriers to inquiry pedagogy in urban classrooms. In Floden, R. (Ed.), *Leadership for Meaningful Learning Using Technology: What Educators Need to Know and Do*, Teachers College Press, New York.

Songer, N. B., Huber, A., and Lee, S. Y. (2003). Evolution of the technology-rich, inquiry-based BioKIDS curriculum. Paper presented at the *Annual Meeting of the National Association of Research in Science Teaching (NARST)*, Philadelphia, PA.

Soreanu, P. (1998). What is wrong with computer-mediated communication (CMC)—An educator's point of view. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, ACM Press, New York. pp. 46–47.

Spain, K. A., Phipps, C. A., Rogers, M. E., and Chapparo, B. S. (2001). Data collection in the Palm of your hand: A case study. *International Journal of Human–Computer Interaction* 13: 231–243.

Thornton, R. K., and Sokoloff, D. R. (1990). Learning motion concepts using real-time microcomputer-based laboratory tools. *American Journal of Physics* 58: 858–870.

Tseng, H. M., Tiplady, B., Macleod, H. A., and Wright, P. (1998). Computer anxiety: A comparison of pen-based personal digital assistants, conventional computer and paper assessment of mood and performance. *British Journal of Psychology* 89: 599–610.

Weber, B. A., and Roberts, B. L. (2000). Data collection using handheld computers. *Nursing Research* 49: 173–175.