CareLog: A Selective Archiving Tool for Behavior Management in Schools

Gillian R. Hayes
Donald Bren School of Information and Computer Sciences
University of California, Irvine
Irvine, California USA
gillianrh@ics.uci.edu

Lamar M. Gardere, Gregory D. Abowd
School of Interactive Computing & GVU Center
Georgia Institute of Technology
Atlanta, Georgia USA
lmgardere@hotmail.com, abowd@cc.gatech.edu

Khai N. Truong
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada
khai@cs.toronto.edu

ABSTRACT
Identifying the function of problem behavior can lead to the development of more effective interventions. One way to identify the function is through functional behavior assessment (FBA). Teachers conduct FBA in schools. However, the task load of recording the data manually is high, and the challenge of accurately identifying antecedents and consequences is significant while interacting with students. These issues often result in imperfect information capture. CareLog allows teachers more easily to conduct FBAs and enhances the capture of relevant information. In this paper, we describe the design process that led to five design principles that governed the development of CareLog. We present results from a five-month, quasi-controlled study aimed at validating those design principles. We reflect on how various constraints imposed by special education settings impact the design and evaluation process for HCI practitioners and researchers.

Author Keywords
Capture and access, special education, selective archiving.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous, K4.2 :Computers and Society: Social Issues, J.3 Computer Applications: Life and Medical Sciences

BACKGROUND AND INTRODUCTION

Management of severe problem behaviors for children has become a primary focus in many classrooms [1]. Such behaviors can disrupt the educational activities for the student with the behavior disorder and for the other students in the school. Furthermore, many students exhibiting these severe behaviors also have other disabilities, frequently Autism Spectrum Disorder (ASD) but also Down's Syndrome, Fragile X, and other developmental disorders. These disabilities can make traditional behavior management and teaching techniques difficult if not impossible.

Behavioral intervention plans based on an understanding of "why" a particular child exhibits detrimental behavior can be extremely useful in addressing a wide range of problem behaviors, particularly for children with severe behavioral disabilities. Functional Behavior Assessment (FBA) is a commonly employed technique for understanding and addressing problem behavior [10]. It relies on a variety of strategies to identify the functions or purposes of specific behaviors and to help educators, parents, and other advocates develop interventions to address directly the problem behavior. The focus when conducting FBA is on identifying significant, child-specific social, affective, cognitive, and/or environmental factors associated with the occurrence or lack of specific behaviors [12]. This broad perspective offers understanding of the function behind behavior, enabling caregivers to develop specific behavioral interventions. These techniques have been shown to be effective both in clinical and in more natural (*e.g.*, classroom) settings [10, 13, 15]. Typically, they are used in special education classrooms as described in this work, but many experts with whom we worked also noted the need for using such techniques more broadly.

FBA includes first identifying and defining the problem behavior, then reviewing information from various sources. It must include both what is known as indirect data collection (*e.g.,* questionnaires, semi-structured interviews with students, teachers, and others) and direct observation (*e.g.*, observations of students in various settings). These data are generally reviewed in a variety of forms, from anecdotal narrative accounts to scatter plots depicting observed incidents of the behavior. The assessor then carefully examines what has been learned about the behavior and its context. This examination leads to a determination of the function(s) of the behavior, the antecedents that trigger it, and the consequences that

CHI 2008 April 5–10, 2008, Florence, Italy.

maintain[1] it. Finally, a treatment team, often consisting of parents, teachers, and other classroom staff, constructs and implements an intervention plan from the FBA.

In this paper, we describe CareLog and the design process and principles used in its development. Using a mixed method design and evaluation approach, we first established five key design principles specific to the creation of a system to support FBA, a challenging but effective strategy, in schools. We validated these guidelines through the design, development, and ultimate deployment of the CareLog system. We tested the use of CareLog in a special education school for five months. The feedback from this user trial, including in-depth interviews, surveys, and analysis of software logs, sheds light on how teachers can use audio and video data coupled with appropriate, manually tagged metadata to enhance their problem-solving process. A major contribution of this work is the description of the design of a recording technology for special education and its subsequent use in a real-life setting. This design and deployment case study provides a deep understanding of the ways technology can and should be used in sensitive settings, such as special education, in which data capture may be unexpected but essential.

CARELOG DESIGN PROCESS

Context of design

In this work, we were designing a research system for a real-life setting. We were exploring research questions, such as the manner in which teachers and students might react to, adopt, and integrate video recording as well as how these types of technology might sensitively support data collection practices. These questions inherently require answers that can only be found in the "real world." Thus, our designs, to properly probe reactions to, adoption of, and understanding about the use of video capture technologies in support of FBA, had to work in these challenging environments and to support practices that are at least partially engrained in those settings.

For children with severe behavior disorders, the implementation of an intervention to minimize the problem behaviors often has dramatic effects on the child's quality of life. Unfortunately, these children cannot explain what is causing the behavior (*e.g.,* why they are in so much pain, why they are hitting themselves and others). To get this information, behavior specialists carefully document the context surrounding the behavior and conduct FBAs. Unfortunately, an FBA undertaken in a clinical setting (typically an experimental lab or behavior management room separate from the classroom) often lacks ecological validity and can lead caregivers to the wrong conclusions and ineffective interventions. On the other hand, FBA undertaken in natural settings is very disruptive. External observers, documentation by teachers and caregivers, and videotaping using traditional methods are invasive, making caregivers and subjects uncomfortable and often rendering the data unreliable and ineffective.

In addition to the contributions of this solution to the fields of education and behavior management, this sensitive domain problem results in an authentic setting in which to push on interesting issues of HCI research. Such seemingly benign issues as efficiency, usability and utility must be reconsidered in this setting. Simultaneously, we must explore such touchstone issues as surveillance, privacy and control of data, and the empowerment of users to make decisions for and about themselves and those in their charge.

Due to the particular challenges of the domain, it was necessary to adapt traditional user-centered design methods. For example, the students whose behavior would be tracked must always be particularly well protected, but their feelings about the technology, in fact about the intervention process itself, could not be ascertained directly. By definition, the children who would be the subjects of inquiry for FBA, many of whom are on the autism spectrum, have trouble communicating and understanding abstract concepts, such as privacy, security, or protection of data. Thus, we relied on our own experiences through participant observation[2] and those of experts. By using information gleaned from behavioral experts and teachers, in addition to that in the literature and from our experience, we were able to triangulate user needs and concerns. Our approach was useful for understanding the user requirements and contributes to the user-centered design arsenal for researchers working with populations who struggle with communication.

Because we were targeting public schools for the deployment of CareLog, specific requirements of the US public school system had to be considered. For example, the technology could not interfere with the network in the schools in any way without going through a rigorous certification process. Furthermore, deployment of new technologies into these real settings required that we develop them in a more robust way than is typical for many research projects. Finally, our designs had to be flexible in use to accommodate both the specific data capture practices advocated by behavioral specialists as well as the needs of teachers on the front line. Often, these requirements ran parallel to one another if not in open conflict.

Design activities

Previously, we explored the ways in which automated capture technologies might be used to support the education and care of children with autism [7]. Building on the results from this multi-year contextual inquiry, we identified support for behavior management as a primary goal. This work's broad focus necessitated further formative explorations to develop the design principles used in the

[1] *Maintain* encompasses encouraging, continuing to allow, and inadvertently influencing the child to continue the behavior.

[2] All the authors have extensive experience designing and developing technologies for autism; two are trained as behavior therapists.

creation of CareLog. Thus, we used focus groups (28 participants small groups up to 6 individuals) to augment understanding of this domain problem [7]. Finally, we used participatory design with six experts and four teachers, taking advantage of the inclusion of a special education researcher on our design team as well as relationships with FBA experts in three local public schools. Paper prototypes, both high and low fidelity, were useful in iterating quickly on the designs within the team (see for example, Fig 1). Domain experts on the design team often sketched out new ideas on paper from scratch (see for example, Fig 2).

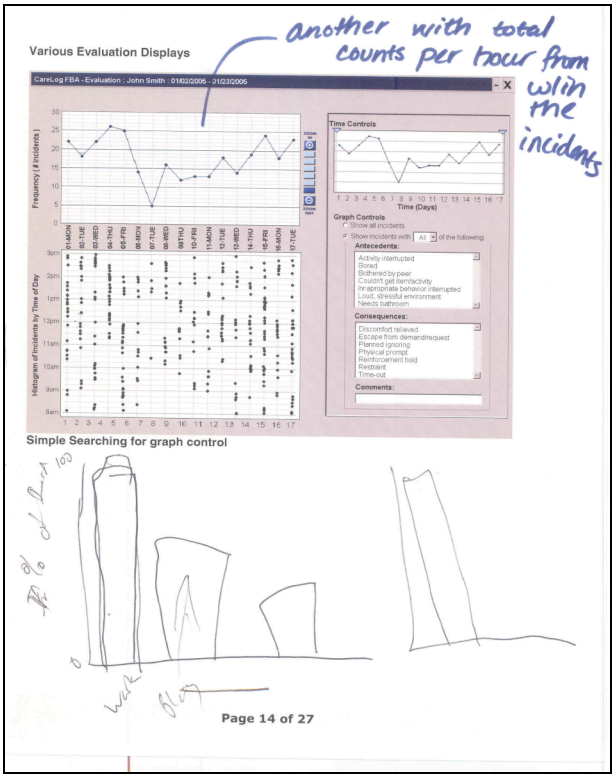

Figure 1: An annotated paper prototype used in a session with a behavior expert working for a local school system.

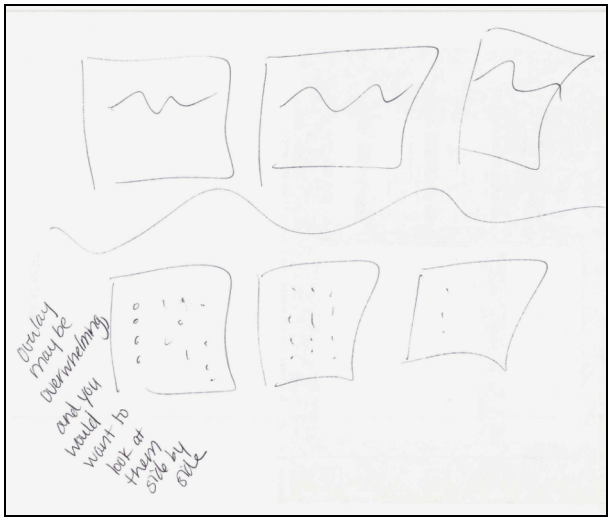

Figure 2. A design sketch created by a behavioral expert as part of a cooperative design session.

DESIGN PRINCIPLES

Using the results from our mixed method formative evaluations [7, 8, 9] and cooperative design sessions, we developed five key design principles specific to the creation of a system to support FBA in schools. In this section, we describe those principles and the ways in which our solution addresses them. We also describe how each principle was implemented in the completed system as it was deployed in one school. The technical details of the system are available in [6].

1. Get the data, "all the data[3]"

Conducting FBA requires the collection of data regarding a large number of behavioral incidents (usually 30 to 50, which can take weeks or years depending on the child) as well as the context surrounding, antecedents to, and consequences of these incidents. At the same time, students may demonstrate these behaviors at unpredictable times, possibly only for a moment, and sometimes with severe enough consequences that distraction and focus of the staff elsewhere is highly likely. The teachers and staff members attempting to understand these complex events often miss the details they need to understand the behaviors. Thus, we needed to supply teachers with the ability to gather these details in a fairly automatic way.

Automated capture and access technologies, as defined by Abowd and Mynatt [2], allow for the constant recording of information from live events, such as audio and video, for successful review at a later time. Similarly, the Smart Kindergarten project focuses on integrating video streams with a variety of automatically sensed data [14]. The most promising feature of these technologies is their ability to offload some of the burden of annotation from the users whose primary activities require their full attention. For example, teachers who may be restraining a violent child or chasing a student who is running away from school cannot take careful data in the moment. However, detailed information is essential in these instances. Automated capture technologies can thereby support caregivers of children with severe behaviors who often cannot manually record the information they need.

2. Empower the people

Despite their potential benefits in the school setting, particularly in situations in which the primary task is particularly challenging, these technologies bring with them concerns about control of data, recording of too much information that is then hard to mine, and other socio-technical tensions [7]. Teachers expressed repeatedly the desire to control recording of data along with other aspects of their classroom activities, such as their behavior management plans and curriculum development. Thus, we sought a solution that would allow for teacher control of

[3] Although it is unreasonable to get all the data about a behavior, experts requested "all the data" meaning all the *incidents*.

archiving and access. At the same time, we offloaded the burden of beginning those recordings at the appropriate time to capture the right information.

Selective archiving uses a collection of capture services embedded in an environment. These services are always on and available but require explicit user action to store an experience. Use of selective archiving in our design offloads a significant portion of the burden of capture while maintaining teacher autonomy. Furthermore, we added the additional safeguard of allowing teachers to delete any data that was erroneously saved using selective archiving when they are reviewing their saved information.

3. Keep it simple, really simple

If teachers are to be in control of data capture, they must be able to do so in an incredibly simple and straightforward manner that blends almost seamlessly into their standard daily activities. The primary task of a teacher or a teacher's aid is, and must be, teaching and caring for the students in the classroom. Although data capture is already a part of those care activities, classroom staff almost never let data capture, which is primarily beneficial in the long term, to interfere with direct educational and care activities, which are more beneficial in the short term. To accommodate this need, CareLog provides teachers with a simple remote interface (see Fig 3), through which a single button press can actuate archiving of data that documents both what occurred in the past and what is to occur in the future. The exact amounts of time to be recorded are set by the classroom staff using a wizard interface that can be reconfigured at any time and is set prior to initial use.

In current practice, behaviorists and teachers both reported that teachers sometimes continue to collect data indefinitely because they never have enough time to pause data collection and assess their data to determine whether further collection is even necessary. Thus, the access interfaces needed to be quick and easy to use. We provided synchronized views of the videos from different cameras to which users could add tags for antecedent, consequence and, context metadata. We provided pre-populated but editable lists of appropriate tags (see Fig 4). The access interface also included analysis tools. The number of incidents, times, and other basic data could be graphed quickly based on just the data created automatically by clicking the button to save video. Teachers could also query the data provided by their tagging activities to support quick evaluations (see Fig 5).

Figure 3: A teacher holds the remote actuator. Each button can be programmed separately, but all participants chose to have all four work identically: either actuate archiving for a set time in the past and future (*e.g.*, five minutes before button press and five after for a total ten minute clip) or actuate a start of recording plus some time in the past and then on the second press, a stop of recording plus some time in the future.

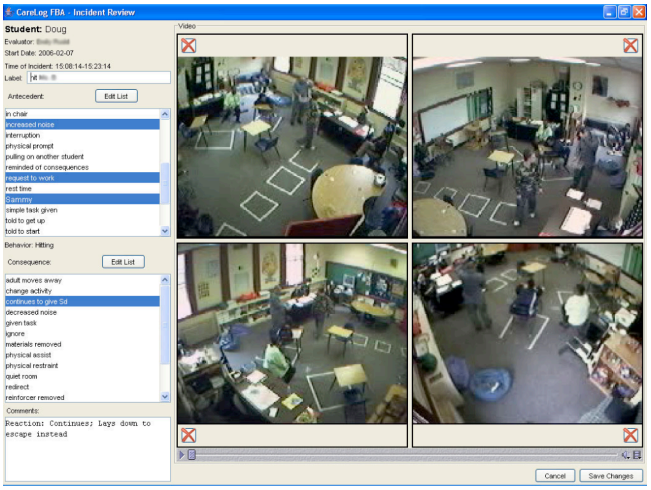

Figure 4: The access interface allows teachers to view all four video streams and one audio feed simultaneously and provide the metadata tags required for FBA (list of tags on left).

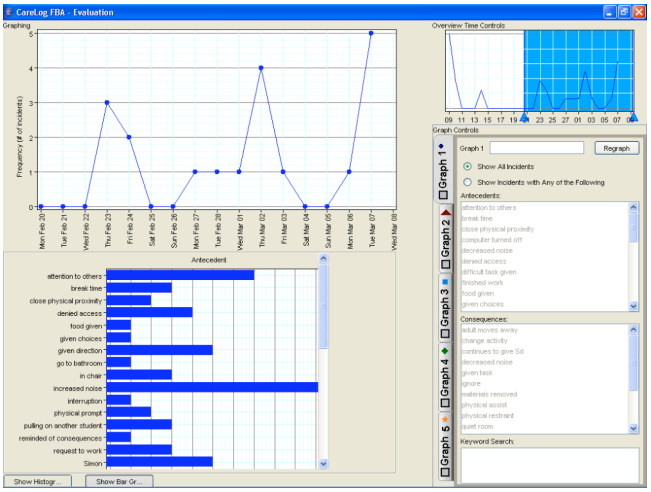

Figure 5: Users are able to view automatically generated graphs that allow for the testing of multiple hypotheses.

4. Shift the burden

During this study, we further unpacked the common complaint by teachers that they lack the time to conduct the data assessment portion of FBA: many teachers do in fact have both the time and the interest to do this work, just not while children and other classroom staff members are present. They spend time teaching, managing behavior of students, and managing personnel, both permanent staff, such as aides, and itinerant staff (*e.g.*, speech therapists). With all of this activity, they are lucky to be able to record the data much less analyze it in depth.

Thus, a primary goal of the CareLog design was to allow teachers to access their data at a time and in a place appropriate for their needs. Using a system in which video is quickly and easily captured for later analysis supports some aspects of this need. We noticed in the first few weeks of our deployment of the technology, however, that the notion of mobility needed to be added to the design. Teachers requested the ability to use the system in a location that was more comfortable and quieter than their classrooms. Furthermore, they had difficulty finding time away from those who should not have access to the data to watch the videos. We developed a mechanism by which teachers could synchronize the data on a laptop with the classroom system. In this way, they were able to take the data home with them or to another quiet location at which inappropriate individuals were rarely present.

5. Reveal and conceal technology safely

The physicality of a classroom environment can be fundamental to the larger concept of an instructional environment. Teachers carefully plan the physical layout and decorations. In special education classrooms, this requirement can be even more significant. For example, all of the teachers with whom we interacted used physical space barriers, such as masking tape on the floor, to denote for students the proper placements of their desks. Any changes to this environment can damage the carefully constructed routines that teachers impart on their classrooms. Deviations may also be extremely distracting to the students. In particular, in the case of autism, children might react in extreme ways to any new unknown stimulus.

In designing the hardware and installation plans for CareLog, we paid particular attention to keeping the classrooms as close to their original states as possible. The installation included four small web cams installed in the ceiling near the four corners of each classroom (see Fig 6). A small pen microphone was installed in the ceiling near the center of the room. A flat-screen monitor with wireless keyboard, mouse, and headphones for interacting with CareLog was included and typically placed on, near, or behind the teacher's desk. Finally, the desktop PCs required to run the system as well as networking and sensor equipment were all stacked on top of a bookcase or other high surface along a wall of each room. This installation plan both minimized distractions and reduced the likelihood that students could reach and damage the equipment.

At the same time, we recognized the need for revealing system status, in particular making visible recording to teachers, staff, and other relevant stakeholders. All cameras had small red lights when powered. Any time an incident was recorded, a small persistent notification and status window on the monitor would indicate that the event had been received and was processing or completed.

EVALUATION METHOD

Designing and developing a data capture system for special education within public education settings also

Figure 6: (left) One of the four cameras in a classroom, all attached to the ceiling unobtrusively. (right) Our computers were stacked on top of tall filing cabinets in each room in the middle of other equipment and boxes typically stored there.

created some challenges to the evaluation process. Important users and stakeholders of this system include teachers, in-classroom staff (*e.g.,* teacher's aides), out of classroom staff (*e.g.,* speech therapists, school administration), parents, and the children themselves. These different stakeholders often have different and sometimes conflicting needs and reactions to new technologies. To make matters more complicated, most of the children involved were *nonverbal*[4]. Some of the students used pictures and other means to communicate with caregivers about simple concepts, such as what to eat for lunch. Feelings about intrusion into the classroom experience, concerns about privacy, and other issues of importance, however, were extremely hard to probe with these children. Even with the teachers and staff themselves, it could be difficult to evaluate the system's true effect given the many potentially confounding variables: the socio-political climate of the school, our relationships with teachers and staff built over weeks and months and even years in some cases of working with them, and so forth. Given these challenges, our evaluation included a mix of quantitative and qualitative, subjective and objective, and observed and reported metrics. In this way, we were able more deeply to understand the results of the deployment of the technology. Furthermore, we placed particular emphasis on the teacher experience while still reflecting upon the needs and responses of all those involved, a tactic that is common in special education. Our overall approach to evaluation includes deploying the technology for an extended time, measuring impact using established quantitative scales, and obtaining a deep understanding of this impact using qualitative, contextualized inquiries into the practices of users involved. By following this approach, others may benefit from our experience in such difficult deployment and evaluation situations in the future.

To fully understand the ways in which CareLog meets the design principles outlined above and identified through

[4] Nonverbal is a domain term meaning they could not communicate through verbal expression.

years of formative work prior to this study, we deployed and evaluated this technology in a real school setting. Four teachers with minimal experience with FBA at one special school for behavior disabilities each used both traditional (pen and paper) and technology-enhanced methods for conducting FBA's as part of a quasi-controlled study. We chose teachers with minimal experience to ensure that their experience with pen and paper methods were comparable to experience with CareLog. Since this study, experts have also used the system with similar experiences. The study design was a mix of within and between subjects designs in that each teacher experienced both experimental FBA conditions (within), but each student was only a subject of one FBA (between). The conditions were counterbalanced, with two teachers using the technology-enhanced FBA process first and two using the traditional method first. Teachers were randomly assigned to groups, and thus the students were randomly assigned to treatment conditions.

Prior to conducting any FBAs, all teachers completed a 5 hour in-service training that included:

- three and a half hours on the process and scientific methods involved in conducting FBAs;
- one hour devoted specifically to use of CareLog for the technology-enhanced condition; and
- thirty minutes on the questionnaires, daily forms, and interviews that would be required as part of the study.

All participants were at a single site, an in-center facility for children with dual disabilities (DD), autism, and severe emotional and behavioral disorder (SEBD). This site includes behavior specialists as part of a program called Technical Assistance for Severe Behavior. These specialists often conduct FBAs in the classrooms, assist the teachers with intervention plans, and work with children directly on behavior management both within and outside of the classroom setting. This school provided an opportunity for us to work with teachers with a range of experience from first year in this setting to 3 years (all had worked in special education for a minimum of a year outside this setting). Furthermore, we had access to small classrooms (4 to 8 students each) with students who all had diagnoses with cognitive and behavioral disabilities, thus making the environment more amenable to behavior modification, and parents and guardians generally supportive of research. In a regular education classroom where the motivation for recording is not as high, we hypothesize that more challenges in getting buy-in would be met.

As part of their training day, teachers each identified two children in their rooms with severe behavior (1 for each condition per teacher for a total of 8 students). One student exhibited such a low frequency of behavior during the first five weeks of his involvement in the study that he was removed and a new student identified for that classroom making the total number of student participants 9. 7 of the 9 students were diagnosed with autism; most also had other diagnoses. 6 of 9 students were nonverbal. One student passed away during study[5]. Enough behavioral incidents had been collected at the time of her death for the teacher to conduct the assessment and complete a final report.

A variety of instruments were used to evaluate the effects of CareLog on teachers, students, the FBA process, and the environment. We used survey instruments to gather background information as well as to measure perceptions of the teachers about technology, video recording, other school records, privacy, and information security. Each teacher completed three surveys, allowing us to measure any changes between the baseline collected during training and after use of each process. The NASA Task Load Index (NASA-TLX) [5] was used to assess workload of each method of FBA. Each teacher also participated in at least one semi-structured interview per assessment and a final interview at the end of the study (minimally three interviews per teacher), typically lasting between thirty minutes and one hour. All interviews were transcribed and coded by the research team. Finally, all of the assessments prepared by the teachers were experimentally verified in a controlled setting by trained, specialized personnel.

The study was conducted over a five-month period. During that time, 109 days were actively spent collecting data for the assessments.[6] Each teacher spent an average of nearly 30 school days (approximately six weeks) as part of the study, with an average of 13.6 days spent per student on the data collection portion of an FBA (min = 7, max = 20). During this time, teachers collected data on a large number of incidents. In both the pen and paper and CareLog treatments, they chose to delete or disregard portions of this data for various reasons (*e.g.,* erroneous "clicking" and actuating of the system, incomprehensible handwriting on the paper forms, etc.). Ultimately, 245 incidents were used as part of their assessments, with an average of 30.6 incidents per student assessment (min= 15, max = 64).

EVALUATION RESULTS

The long-term deployment of the CareLog system allowed us to test the validity and success of the five design principles as well as the general impact and utility of the technology on the FBA process in schools. A primary goal of this research has been to ensure that teachers who have minimal experience with FBA can, in fact conduct them in their classrooms successfully. The fundamental measure of success for an FBA is traditionally whether the resultant function determined by the teacher can be verified in an experimental setting. We conducted experimental verifications for seven of the eight students. The eighth student passed away during data collection, and so no clinical verification could be completed after the

[5] The child's death was reported to the Institutional Review Board at the Georgia Institute of Technology and ruled unrelated to the study.

[6] Some days are double or triple counted in this number, because different teachers assessed multiple children simultaneously.

assessment. Of the seven remaining students, five were fully verified,[7] and in the sixth case, we were able to verify clinically two of the three functions hypothesized by the teacher, but the third was not seen in the clinical setting. The final student's assessment could not be verified, because he did not demonstrate the problem behavior in the clinical setting under any conditions. Long-term assessment of the quality of the treatment is currently being tracked, but expert review of the intervention plans and assessments completed by the teachers indicates that they are in keeping with current best practices.

Inability to verify the function of behavior in a clinical setting does not mean the teacher was wrong. In the unverified case in this study, the teacher attributed the function to attention, particularly attention from certain staff members. These conditions are difficult, if not impossible, to replicate outside of a classroom setting, thus exemplifying one of the problems with clinical analyses that have driven many specialists to focus on natural settings, like the classroom or the home.

Once assured that teachers could conduct successful assessments, it was also important to understand the ways in which the designs, formed from the five guiding principles, impacted the daily lives of the teachers. In the following subsections, we describe the ways in which CareLog successfully addressed the design issues.

1. Automated capture can ensure getting all the data

In the traditional pen and paper method, caregivers and behavior specialists were often restricted in their assessments by what they could witness and document. Many incidents can be missed using this method, resulting in records that are at best incomplete. These incomplete records can lead to development of interventions that do not work and must be redeveloped after months of no progress.

To establish ground truth data on number of behavioral incidents, a member of the research staff recorded an average of 17½ hours or 21.5% of time in study for each student with a handheld video camera. Independent coders, hired and trained specifically for this project, then coded the videos for behavioral incidents using operational definitions created by the research team and teachers. At least two video coders worked on each student's recordings. To ensure reliable data coding and inter-rater agreement, all coders first analyzed a single subset of data, at least one day per every five the student was recorded. Once at least 80% agreement was obtained for these data, the coders continued independently.

We tested whether there was any change in the error rate observed, that is whether there was any difference in the rate at which teachers missed incidents or erroneously recorded data when no incident had occurred. Error rate was calculated from the baseline data observed by the independent data coders and the records created by the teachers that were used in the assessments. Across all of the incidents, only one false positive (recording of an incident when none occurred) was observed, and thus, we do not comment on false positives further in this paper. In this study, teachers made an average of 43.37% fewer false negative errors (missed incidents, from here on, simply referenced as errors), $t(6)=4.9807$, $p<0.005$[8] than with the traditional pen and paper method.

Because each teacher conducted two assessments, a possibility existed that the improvement was due to which order the teacher conducted the assessments. Thus, we also measured the difference between the first and the second assessments. We observed an overall trend towards an increase in error (26%) $t(6)=0.9264$, $p<0.5$, between the first assessment group and the second.

Several reasons for the improvement in error rate with use of CareLog emerged during teacher interviews. Teachers commented that they were able to use CareLog to gather data even when not present. For example, one noted that she missed fewer incidents in her room because

> *"if I'm just like going to get lunch or whatever, a lot of times, … I'm not gone longer than five minutes, so I can go, and if they [her staff] say 'oh he hit' or whatever, I can just press the button and know that it will be in that five minutes."*

Teachers also commented on their ability to see what other students were doing when the staff is occupied with the FBA subject, as exemplified in the following excerpt:

> Teacher: *I also remember the other time I was out of the room, it was interesting to see what was going on with the other students. Because there was an incident that was not caught while the other incident was being dealt with…So it was very interesting to see who got away with what, you know, when …*
> Interviewer: *Because there was less staff…?*
> Teacher: *Right, exactly…Yeah, so and not intentionally, but …*
> Interviewer: *…they sort of know when you're looking and when you're not.*
> Teacher: *Yeah, exactly, so it was funny cause I had to watch it like three or four times, because I was like "Nu-uh" (laughs) and then you know, I would rewind it and be like "Yeah, what a sneaky devil" (laughs)*

In an effort to ensure that teachers could record as many behavioral incidents as possible, we also explored the optimal camera placement and the optimal number of cameras required for teachers. Although formative designs indicated that multiple cameras was likely desirable, we were concerned that the sheer quantities of data generated might be more than is needed when deployed in a school. All of the teacher participants, however, universally reported that having all four camera angles in their

[7] Full verification means the problem behavior was produced reliably in experimental conditions designed to test the possible functions.

[8] All t-tests are two-tailed, paired.

classrooms was a key contributor to their abilities to understand what was happening. They were able to switch between attending to the different angles being shown on the access interface as necessary after only a few attempts. The low resolution of video (320 X 240) we provided was sufficient in all but one of the rooms. This room is usually less well lit than the other rooms, causing some slight viewing problems that would be easily remedied with minor adjustments to the video capture modules.

2. Empowered people benefit from their own actions

When dealing with sensitive situations, such as special education classrooms, and rich detailed data, like audio and video recordings, it is important to understand who is empowered to control that data and how. In our formative explorations for the design of CareLog, we discovered that teachers were likely to adopt openly only those recording technologies over which they have control and in which they have confidence of security. During this study, most of the school staff embraced the technology, weaving it into the social and political fabric of the school. Furthermore, all of the staff still present at the school after the study ended, even those who had been hesitant initially, requested to be part of ongoing studies and to use CareLog more in the future, often commenting on the system's "respect" for current practice and the "trust" they had in it and in us.

Teachers also reported specific benefits of using CareLog to quality of life in the classroom. All of the teachers reported that these benefits and changes in their practices were enabled because they felt "comfortable" and "safe" with the technology and did not view it as a "threat." Primarily, teachers reported that use of the captured videos allowed them greater access to activities in their classrooms, whether staff- or student-related. For example, three out of four of the teachers reported noticing their own teaching style and behavior or that of their staff members. One teacher noted that she could see how well her staff was "following the plan" in terms of behavior intervention and intended to praise them for activities while she was not present.

Teachers also reported being made more aware of their own mistakes while teaching as result of watching the videos provided by CareLog. This type of reflection was not an explicit goal of this system. However, in other work, special education supervisors of classroom staff predicted that they may be able to self-correct those actions that may in fact exacerbate inappropriate behavior when viewing captured videos [7]. Furthermore, in other educational situations, researchers have demonstrated that teachers do self-correct and critique their own actions [3, 4]. Teachers noted that access to these videos was helpful to them, because they were reviewing data that they themselves captured and that no one else could access. These restrictions provided the comfort needed to assess their own performance critically.

3. Keeping it simple enables complex assessment with minimal work

FBAs, although valuable, are often not used in classroom settings because the workload on the part of the teacher (and sometimes the teacher's aides) is simply too high to maintain good records regularly, do the analysis, and perform at a high level in terms of instruction and classroom management. We observed these occurrences within our own fieldwork and understood this problem to be widespread from expert collaborators. Thus, one of the primary goals of this effort was to develop a solution that would reduce the workload of the staff members involved in the FBA process. Use of CareLog substantially reduced reported workload over the pen and paper method both directly in terms of the perceived load of specific tasks and indirectly by redistributing the work to other times or staff members. The direct reduction is addressed in the following paragraphs with the latter result discussed in the next subsection.

Workload is a multi-dimensional psychological construct measuring the subjective experience of work that results from the mental actions performed while perceiving and processing information and executing a response. The NASA-TLX measures workload on 6 different dimensions (Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, Frustration) to measure the amount and type of workload a user experiences during task performance [5]. Use of these sub-scales allows different tasks to be compared for overall complexity for a person, in this case a teacher, to complete. We used it to understand the workload differences between the pen and paper case and the experimental case of using CareLog to conduct FBAs.

We adapted the NASA-TLX scales to suit the way the data was collected. Each teacher was asked to rate the workload of information capture at the end of each day, rather than immediately after each data collection moment (behavioral incident). They were then asked to complete a second workload assessment form at the end of each day that they tagged any data, that is added antecedent, consequence, or context data. Finally, they also completed an assessment form for each day that they conducted any data analysis work (typically only one or two days). This adaptation ensured that teachers were able to complete the rating without significantly disrupting their work and that we were able to examine workload across the entire FBA process or as one of three sub-activities: initial data collection, tagging with metadata, and analysis. Teachers completed as few as one or as many as twenty-two rating forms for any given sub-activity depending on the number of days they performed the activity and their own compliance rates with the experimental protocol. All of the calculations were normalized for a standard 100-point scale regardless of the number of rating forms completed per teacher per activity.

Clicking a button to note that an incident has occurred is unsurprisingly much simpler than writing down the same information on a form or piece of paper. Using CareLog was reported to be significantly easier in terms of workload than the traditional method for recording basic behavioral information ($t(6)=3.8983$, $p<0.01$).

Across all four teachers, we also observed the trend of less workload for determining the function itself. Although this trend did not show significant ($t(6)=1.702$, $p<0.15$), the trend fits with our predictions in that the software itself inherently provides benefits such as automatic organization of information and graphing of data. In future designs, based on this data and feedback obtained during this study, we are including such features as automatic report generation to reduce this workload even further.

Teachers reported no difference in the workload of labeling behavioral incidents with metadata with CareLog versus the traditional method. Our in-depth interviews with teachers indicate that they viewed the work of labeling the data in each of the conditions to be of similar effort levels for different reasons. In the traditional method, they must create these labels at the time of the incident's occurrence—often a very trying time in the classroom—or remember the details later when they have time to make the notes—sometimes a very stressful task. On the other hand, in the technology-enhanced case, this portion of the process requires watching extensive video clips and making determinations about what is happening during those clips.

4. Shifting the burden allows caregivers to work effectively

During the formative design process for CareLog, the numerous required tasks and activities for a teacher in a single day at school came up time and again. Many experts, teachers, and classroom staff commented on the lack of control and spare time special education teachers and staff have over their days. Thus, we designed CareLog to allow users to temporally and physically shift the burden of assessment to a time and place in which they were relaxed, comfortable, and able to concentrate.

Using CareLog, work could be redistributed to times more convenient for the teachers. That is, they could do the labeling of antecedents and consequences after the students had left for the day or over a weekend, as opposed to the requirement of labeling data during or immediately after the incident as with pen and paper. For example, one teacher noted "it was a lot easier to go back when I had time to go back and actually see exactly what was going on." Even though watching the videos took a lot of time and might seem overly laborious, the teachers unanimously reported preferring it, because they could do it when they had time and energy rather than close to the sometimes dramatic and always difficult behavioral incidents as is required in the pen and paper method. Even more compelling, determining the correct function the first time means that the teachers were not required to conduct subsequent assessments after implementing interventions that ultimately failed.

The teachers also unanimously reported being more likely to entrust other staff members to record data when they were out of the room than with the pen and paper method. For example, when asked about the possibility of having staff members take traditional data, one teacher responded:

> *"I would want the consistency of somebody's who's been trained with an FBA to do the antecedents and the consequences [with the pen and paper method]. And that's a big change, between that and the clicker [remote that comes with CareLog] The clicker I could trust with a total stranger, a complete sub, I would expect that they could handle 'click something.' ... Click it ... if you see this, click it. You know, if you just show them one example of the behavior, if they knew what they were clicking on, that was nice. That would be a definite benefit."*

Most teachers did not ask classroom staff to collect data in the pen and paper condition. Only one tried, and that attempt only lasted for one part of one day with both the teacher and the aide noting that the task was too difficult due to the coordination that was required between the teacher and staff. In the CareLog condition, however, three of the four teachers asked an aide to record data for more than a day. In one classroom, an aide was the primary data collector with the teacher in that classroom completing the labeling and analysis of the data collected by her staff member.

5. Appropriate visibility breeds adoption

A primary consideration of this work was to minimize the impact on the physical environment and thus the obtrusiveness of the technology to the classroom environment and students. The placement of small cameras in the ceiling succeeded in minimizing the amount of student distraction. Only one student in any of the rooms (total of 22 students) noticed the overhead cameras. Every student who could communicate verbally, however, noticed and commented on the camera carried by research staff when ground truth baseline data was being collected using a handheld camera. It is an open policy question for teachers and administrators how communication about these systems should be handled in non-research deployments.

The invisibility of the system to the students in these classrooms was often cited as a highly desirable feature for continued use of the system. One of the biggest struggles reported by teachers in the formative design stages of this work and throughout the deployment study centered on recording data to share with others or to review themselves. They wanted to include behavior that was "natural," particularly for those behaviors that are "quirky." Often these difficult behaviors can be hard to reproduce outside of the classroom. Furthermore, they can even be hard to reproduce in the classroom when that environment has changed (*e.g.,* with the addition of an external observer).

Unfortunately, the system was also occasionally invisible to other staff who would come into an experimental classroom. Although these individuals were briefed at the beginning of the study about the recording in various classrooms, one teacher noted that two staff members not directly participating in the study commented that they had forgotten the recording was taking place and had worried that something undesirable might have been captured. The

concerns abated when the teacher explained that she had not chosen to archive anything, and thus, any recordings had been automatically deleted by the system.

Many school specialists argue that eventually all schools will have complete video infrastructures and therefore, an assumption will be made that when one is in school, one is also being recorded [7]. In the short-term, however, researchers and schools implementing these systems must consider the broad implications for the school as a whole and the balance of invisibility and the impact of systems on their environments and users with the need or desire for continual informed consent.

CONCLUSIONS AND FUTURE WORK

Behavior management, a significant goal for special education, can be augmented greatly by better understanding of the *functions* of inappropriate behaviors. Educators and behavioral experts use the best practice of FBA to determine these functions. This process is difficult, time-consuming and error-prone, particularly when practiced in the natural environments of classrooms and homes.

We developed an application to support FBA called CareLog and evaluated it in a deployment study with four teachers and eight students in one special education school. Using a mixed method approach to understand the design considerations for technology to support FBA, we developed five key design principles for the creation of a system to support FBA in schools. We then used these guidelines to design, develop, and deploy CareLog. This case study demonstrates the usage of the design guidelines produces a successful system for conducting FBAs in schools. Our evaluation shows the system can be usable by teachers with minimal training and little impact on their workload. Despite the potential danger of being an invasive application, the system resulted in little intrusion to the environment and the teaching activity.

Too often iterative design is only used at the start of the design phase and stops at the deployment stage. Beyond the research contributions stated above, this case study also demonstrates the importance of continuing the iterative design process throughout the research project. As a research prototype, the system had a few errors which were minimal and eliminated by the end of the study. Although there was room for improvements with respect to how we designed, deployed and evaluated, we benefited from sympathetic and highly motivated participants. Furthermore, the system was far more successful than our evaluation results can reveal. After completing our evaluation, the system remains in use. One of the teachers in this study is continuing to use our research prototype on a long-term basis, and other teachers in the same school have requested their own systems when they become available. Other teachers and behavior specialists have also requested to use CareLog (either from the study or from hearing about the work). We are currently working to develop a system robust enough to deploy across multiple sites and numerous classrooms.

ACKNOWLEDGMENTS

Thanks to the teachers, school staff, parents, and students who participated in this work; to Juane Heflin and Carina DeFazio, your assistance in the design and deployment of this system was invaluable; to Celeste Buckhalter, Stephen Hosaflook, Sharjeel Hooda, Ellen Matthews, Nicki Pirouz, Jesslynn Beattie, and Jahmeilah Richardson for help in data and software coding; to Caring Technologies, Intel Research, IBM Research, and Google for financial support; to Julie Kientz, Shwetak Patel, Jay Summet, and Trevor Pering for other support and early edits of this manuscript

REFERENCES

1. Education Development Center: Supporting Children with Challenging Behaviors: Relationships Are Key. *Training Guides for the Head Start Learning Community.* Administration for Children, Youth, and Families (DHHS). (1997), 120.
2. Abowd, G.D. and Mynatt, E.D. Charting past, present and future research in ubiquitous computing. *Transactions on Computer-Human Interaction (TOCHI)*, (2000) **7**(1): 29-58.
3. Brophy, J. (Ed.) (2004). Advances in research on teaching: Vol. 10. Using video in teacher education. Oxford, UK: Elsevier.
4. Clarke, D., & Hollingsworth, H. (2000). Seeing is understanding. *Journal of Staff Development*, 21(4), 40-43.
5. Hart, S.G. and L.E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in *Human Mental Workload*, P.A. Hancock and N. Mashkati, Eds. Elsevier (1988) 139-183.
6. Hayes, G.R. *Documenting and Understanding Everyday Activities through the Selective Archiving of Live Experiences*. PhD Thesis. Georgia Tech (2007).
7. Hayes, G.R. and Abowd, G.D. Tensions in Designing Capture Technologies for an Evidence-Based Care Community, in *Proc. CHI 2006*, ACM Press (2006) 937-946.
8. Hayes, G.R. *et al.*, Designing Capture Applications to Support the Education of Children with Autism, in *Proc. Ubicomp 2004,* Springer-Verlag (2004).
9. Hayes, G.R., *et al.*, Experience buffers: a socially appropriate, selective archiving tool for evidence-based care, in *CHI '05 Extended Abstracts,* ACM Press (2005), 1435 - 1438.
10. Horner, R.H. and Carr, E.G. Behavioral Support for Students with Severe Disabilities: Functional Assessment and Comprehensive Intervention. *Journal of Special Education* (1997). **31**(1): 84-104.
11. Iwata, B.A. *et al.*, Toward a functional analysis of self injury. *Analysis and Intervention in Developmental Disabilities* (1982).
12. Quinn, M. M., *et al.* (1998). Addressing student problem behavior: An IEP team's introduction to functional behavioral assessment and behavior intervention plans (2nd ed.). Washington, DC: Center for Effective Collaboration and Practice.
13. Sasso, G.M. *et al.*, Use of descriptive and experimental analyses to identify the functional properties of aberrant behavior in school settings. *Journal of Applied Behavior Analysis* (1992) **25**.
14. Su, N. *et al.*, Augmenting Film and Video Footage with Sensor Data, in *Proc. of PerCom '04*. IEEE Press (2004), 3-12.
15. Umbreit, J., Functional Assessment and Intervention in a Regular Classroom Setting for the Disruptive Behavior of a Student with Attention Deficit Hyperactivity Disorder. *Behavioral Disorders,* (1997) **20**(4): 267-278.