Competency 3: Research and Assessment
Lessons in the Conduct of Effective Evaluations

Kajal Madeka

Introduction

In a world where educational experiences are increasingly mediated through technology and everything from infant cognitive development to advanced adult learning programs come to us via webpages and apps, the question arises, how good are these in serving the purpose for which they were designed? This and many such questions on performance, worth, outcomes, failures etc. puts program evaluation in a central place as a set of methodologies capable of providing answers to such questions. This essay combines my experiences in two evaluation settings–one a traditional class-room based usability project that taught me the ropes of program evaluation and another, an avant-garde undertaking where a dedicated space was created to carry out a variety of evaluation studies and research on cutting-edge educational technology. The essay highlights how these two experiences contributed to my understanding of 'user-centered' testing design in the application of technology to conduct effective assessment and research.

Test of an Inductive Reasoning Prototype: rising to the challenges of working with children.

The instructional technology program that formed the object of evaluation was a prototype of an online inductive reasoning game meant for children from grades four to six, to help them acquire inductive, proportional reasoning and problem solving skills to aid in the transition to Algebra. It was the application of Dr. Gary Phye's research on cognitive training for children and knowledge transfer (Klauger & Phye, 1994). The questions that Dr. Phye and the game designer wanted answered through the evaluation were about general usability with the target audience in mind such as ease of navigation and use, appropriateness of vocabulary used for target audience, visual design and satisfaction.

Unorthodox Test Strategies

Since the test participants were elementary school children, our team conducted some research on how best to design and conduct the test (what evaluative instruments, protocols, methods of data collection to use) and how to modify these for use with children to yield maximum information on the desired questions. Literature on this subject was not easy to come by and our text books did not cover this angle at all. Three internet publications that helped ground the test design in some practical research were Hanna, Risden, & Alexander, 1997; Nielsen, 2002; Dunn, 2006.

Based on these readings, we decided to use specific performance-based tasks to direct the usability testing as elementary school age children are accustomed to following directions and performing tasks given by an adult. A detailed observation grid was designed to record not only the task performance but also associated actions and behaviors. It was acknowledged that children may be hesitant to express themselves truthfully to the adult for fear of displeasing them, but the non-verbal cues and behaviors such as fidgeting, yawns, sighs, frowns, smiles, posture etc. can potentially reveal unspoken thoughts and dispositions. To capture these important pieces of data, the observations were supplemented by video and sound recordings of the entire test. The program database recorded each keystroke and page visited, actions taken and time spent. But the most

effective modification was made to the 'think-aloud' protocol where the participant(s) are asked to verbalize their feelings and thoughts as they performed the tasks. This technique of testing is very powerful in giving insights into how the participant is conceptually interacting with the software (Rubin, 1994) but is challenging to administer with children, raising concerns about articulation and distractedness. To gain the depth of information this measure gives, an innovative approach was adopted wherein the test was conducted with the children working in pairs and while doing so, were encouraged to talk to each other to work out any problems they were facing. We hypothesized that by simulating classroom small-group work, the children would feel more at ease with the testing environment. Children's recall of the usability experience is usually in general terms of the overall impression it left on them (e.g. good, fun, cool), even if parts of it had been frustrating (Hanna et al., 1997). Thus, the debriefing was also executed a little differently by walking through different parts of the program with the child participants and asking questions to encourage specific recall. In this way, by using a combination of technology for data collection and thoughtful, informed modifications of testing norms, not only were we able to collect and cross reference very rich data but also get valuable feedback from the participants that helped identify several areas of change in the program.

My main take-away from this experience was that for effective evaluations to happen, it is essential to tailor the test design and set up to the characteristics of the participants and users involved. Looking back on the inductive reasoning prototype test now, I couldn't help thinking how much easier it would have been to conduct this test in the usability lab with its advanced testing and data collection technology allowing for an unobtrusive setup (no video cameras and wires trailing across the room), in-depth observation and greater breadth of analysis.

Vision of the Usability Lab

The Usability Lab located in N058 Lagomarcino Hall was set up in the Fall of 2008 with funding from the Computation Advisory Committee (CAC) and operated as part of the Center for Technology in Learning and Teaching (CTLT). The lab was established with the purpose of providing students across campus a permanent space and means of scientific data collection to carry out systematic usability studies and evaluation activities to test the efficacy of their own technological solutions or to conduct research (CAC Central Pool Proposal, 2008-2009).

Equipment and Setup

Accordingly, the lab was equipped with the latest hardware and software systems for complete and comprehensive audio and video recording, screen capture and analysis of usability tests. The initial setup of the lab was conceived along the 'electronic observation room' model that is recommended for organizations with limited resources and are new to the field of usability testing (Rubin, 1994). In this model, the testing and observations rooms are physically separate and computer and audio feeds are routed electronically to the observation room where observers can watch ongoing tests on a computer monitor. Two-way communication is not usually possible in such a set up. Thus two rooms: N058 and its annex N058A in Lagomarcino Hall were chosen for the lab. Six workstations were arranged along the periphery of the testing room with a common worktable in the center for activities like introductions and debriefing, for focus group studies, paper-prototype and object usability. An observer station was placed in the observation room and all computers were networked via the CTLT network.

The lab evidently seemed to fill a need because students were making requests to conduct studies even as setup was being completed and equipment being configured. During my one year as Lab Assistant, a variety of different usability activities were conducted in the lab by students from across campus. Some of these included web-site usability study, usability study of the XO computer, computer interface design prototype testing, paper-prototype study, camera usability, learning evaluation of a web-based game on learning of color concepts and comparative study of user experience for Mexican designed vs. American designed travel websites. Apart from this, dissemination and training activities were also conducted in the lab with students of usability.

"Usability" of the Usability Lab

The different focus of every study as well as the needs and number of the participants involved demanded different testing scenarios that enabled me, as the point person, to experiment with a range of test setups, providing valuable experience and data for improvements to the testing space and equipment. It also revealed the limitations of the proposed design and equipment set up vis-à-vis the implementers and target users. For instance, the suggested setup and list of technology presumed a very formal and professional setup with mounting and hardwiring of sound and camera feeds to the observation room and into storage devices, but that required some rewiring and contract work that wasn't discussed before or budgeted causing some reluctance for implementation. There were some other items on the list whose purpose was not made very clear to the staff in charge of the setup and consequently were dropped from the purchase list to achieve some cost saving. Similarly, the users (students) sought more flexibility in the conduct of the tests. From the first test conducted in the lab, it became clear that the students were very uncomfortable with being in the observation room separated from the participant, especially if they were children. This was because unlike a professional setup where there are a team of people fulfilling different testing roles, the students doubled in their roles as observer and test administrator and found themselves shuttling between the two rooms as the test progressed. They preferred to be in physical proximity to interact with the test participant if needed and often chose to use one of the other workstations to record and observe (Figure 3).

There were other challenges too with using the annex as an observation room. The proposed plan seemed to have presumed that annex N058A would be a dedicated observation room, but in reality, it also doubled as the office for two research assistants (I being one of them) and for storage of usability and other hardware, making it quite cramped and busy for quiet observation. Taking all this into consideration, some changes were made to the original layout wherein the observer station was moved out to the testing room where its powerful capabilities would be better utilized and also accommodate the working-style preferences of the students (Figure 2). A little corner niche was formed by placement of tables so that the screen and the person behind it was reasonably shielded from view and could observe unobtrusively, yet be close at hand to the participant.

Thus, I learned that like all innovations, even a lab set up needs to be compatible with the characteristics of the target users, needs to be done in consultation with the key stakeholders for full implementation, and like a well-designed test, needs to be flexible enough to customize.

References

Dunn, T. (July 2006) *Usability testing with children*.. Retrieved 4th March, 2008 and 28th Oct 2014 from http://www.webcredible.co.uk/user-friendly-resources/web-usability/testing-children.shtml

Hanna, L., Risden, K., and Alexander, K. (1997). Guidelines for Usability Testing with Children. *Interactions*, *4*(5). Retrieved 4th March, 2008 and 28th Oct 2014 from http://www.microsoft.com/usability/UEPostings/p9-hanna.pdf

Klauer, K. J. & Phye, G. D. (1994). *Cognitive training for children: A developmental program of inductive reasoning and problem solving*. Seattle, WA: Hogrefe & Huber Publishers.

Nielsen, J. (2002, April 14). Kids' corner: Website usability for children. Retrieved 4th March, 2008 and 28th Oct 2014 from http://www.useit.com/alertbox/children.html

Rubin, J. (1994) *Handbook of Usability Testing.* John Wiley & Sons, Canada.