Use of Computer Interface Usability Testing to Evaluate the Success of Science Communication Website Artifacts.

Final Research Paper
Science Communication Program
Laurentian University
A.J. Boulay G Dip. Candidate
Dr. David Goforth, Supervisor
August 13, 2008

Introduction

The problem that is under study is the problem of improving the communication of science on internet websites. The specific website examined is that of the Center for Rural and Northern Health Research (CRaNHR) at Laurentian University, Sudbury Ontario. This website holds a great deal of scientific writing in the form of published papers, reports, newsletters, unpublished research and links to other sites that also carry pertinent research documents. The question that is addressed in this research is “How does the Usability of the computer system affect how well the science of that website is communicated?” Much of the study of computer system Usability occurs in the marketplace rather than the scientific laboratory. For this reason, the scientific treatment of this issue is sparse. However, the marketplace has provided solid standards that allow one to evaluate Usability in general. These standards are provided by the International Organization for Standardization, and address the design requirements for information technology and multimedia systems with regard to the ergonomics of human system interaction. Usability is the measure of human-system interaction and it is measured through Usability testing. In this study, the Usability of the CRaNHR website is evaluated in regard to the scientific communication contained therein. The standards used in this study to measure the Usability are found under ISO 9241-10 and ISO 9241-11 (1996), Dialogue Principles. Up to now, only a few instruments that support the summative and formative evaluations according to ISO 9241 parts 10 and 11 have been developed. However, there has been much consideration of Usability and human-computer interaction in market driven documents.

Opaluch states that one should “Identify the measurements that are most central to the success of the product or service being evaluated”, (Ratner, 2003, p. 119). Because the product of the health research center is the science that it produces, the success of the CRaNHR website is contingent on effectively communicating the science that is held in the website. Thus, an operational definition of Usability must not only reflect the ISO standards that permeate the medium, but also consider the goal of the website being evaluated. The goals of the study were described by the researchers at CRaNHR and include improvement of the Usability of the website and an improvement of the navigation undertaken within the website. These two criteria are mutually inclusive as an improvement of the navigation undertaken within the website. Thus, the hypothesis and experimental design relate to the problem of communicating the science held on the CRaNHR website with regard to increased Usability and improved navigation.

The theoretical implications of this study are pervasive within the medium of Science Communication. Science Communication specializes in producing communication products that provide successful learning activities in the area of science. The CRaNHR website holds an audience of health professionals that is accustomed to the format and presentation of the complicated scientific information. However, improvements to the Usability of the website will provide persons who do not have a formal understanding of health science with an opportunity to appreciate and enjoy their engagement with CRaNHR's scientific endeavours. Because there is little previous work that examines the communication of science on websites outside of strict educational media, this study cannot be related directly to any existing body of research. However, this study will contribute to the field of Science Communication by examining the relationship between accepted Science Communication theory and existent theories pertinent to human-computer interface. Finally, the results of the experiment administered in this study will provide an indication of the Usability of the CRaNHR website.

There is a strong relation between human-computer interface design and the design of Science Communication artifacts in science museums. Both of these areas exploit the interactive component of the user's experience and so an examination of the interactive components of Science Communication within a website is justified. Specifically, in both areas there is a high regard for treating context according to the interaction between the system and the visitor. In Science Communication design, context is used to produce cognitive avenues that will best promote the understanding of new scientific information. Specifically, authors such as Rochelle (1995), Cziksentmihalyi and Hermanson (1995), Hedge (1995) and Vygotsky, (Hein, 2005) have discussed the importance of context in the areas of Science Communication and informal learning. However, there is little crossover by authors between the two genres; only one paper has been found that cites Jeremy Rochelle in the HCI literature (Russe, David, R., 2001). Because scientific information reflects the kind of reasoning that is inherent in producing science, the development of scientific knowledge generally progresses from a foundation of ‘first principles’ toward the accumulation of associated theories. Thus, the apprehension of new scientific knowledge must avoid any kind of dissonance between previously learned material and the new scientific material to avoid errors in understanding. Avoiding errors may be achieved by attaining a ‘zone of proximal development’, (Vygotsky in Hein, 2005), where the visitor may ‘chunk’ new information with previously learned information and avoid inappropriate cognitive dissonance between new and old understandings. This is the kind of approach that would work best with non-experts who are not familiar with the character of the scientific genre.

In the area of human-computer interface, the structure of the information within the system is essential to how successful the dialogue between the user and the system can be. For example, a web site may choose to organize its information reflective of the processes in the organization. In a research center such as CRaNHR, the organizational character of produces a structure that emphasizes ‘published papers’ and ‘research presentations’ rather than emphasizing the treatment of specific topics across many product media, such as ‘nursing’, across all media. This means that a visitor may apply an inappropriate strategy when beginning to review the contents of the CRaHNR website. Thus, the way that the information is structured in the CRaHNR website interacts with the visitor. This is because the practice of website Usability testing is concerned with evaluating the organization of information within a website. It is submitted that improvements in the Science Communication experience obtained by its visitors. Usability testing is also concerned with evaluating how visitors use the organization of information to successfully complete tasks. Importantly, Usability evaluations will also consider satisfaction levels among website users. The higher the level of satisfaction, the greater the likelihood that a successful Science Communication experience will have occurred.

Organization of information, navigation and task completion and satisfaction are the three main areas of concern for this Usability evaluation. Organization of information was tested using a card sort task that identified natural categories of information. Navigation and task completion was tested using videotaped trials where subjects were given different search tasks and asked to reach prescribed goals. Finally, the effectiveness, efficiency and satisfaction levels of the subjects regarding the website were measured using the System Usability Scale, (SUS), a ten item Likert satisfaction survey as defined in ISO 9241-11. All of these measures are context specific.

This study examines the theoretical proposition that the relationship between effective Science Communication that successfully treats context between the visitor and the website may be measured through human-computer interface (HCI) assessments. The results of the HCI Usability evaluation may positively interact with the success of the Science Communication. This proposition is further strengthened by the fact that Vygotskian Activity Theory is applied in both the Science Communication and genre and the HCI Usability Testing genre. However, Vygotskian Activity theory is only one of the measurement tools that will be used. Other metrics such as those that satisfy the International Standards Organization (ISO) Usability assessments are also used. It is submitted that these metrics of HCI may also provide a basis for the future creation of an instrument that will effectively evaluate Science Communication systems and networks. It is suggested that increased effectiveness, efficiency and satisfaction scores in Usability studies of web site Science Communication will provide data that may be used to improve Science Communication information technology artifacts. However, this is simply a pilot study or a ‘quick and dirty’ analysis that will only provide a foundation for further detailed examination of these questions.

Methods

Participants

Subjects that completed this pilot study ranged in age between 18 yrs and 25 yrs old. The gender distribution was half male and half female. Subjects were selected from a sample of employees of the Science North Science Center, Sudbury Ontario .

The three tests were administered to each of the subjects individually.

Apparatus

The apparatus differed for each of the three tests.

The first test was the Card Sort Test where subjects were asked to sort index cards. The index cards each had a different label and each label corresponded to a major topic in the CRaHNR website. The selection of each of the topic labels followed a method of content analysis specific to a semantic network approach (Krippendorff, 2005, pp. 292-288). This method of content analysis chosen to extract the major keywords from the CRaHNR website utilized theory specific to computational modelling of language cognition (McLeod, Plunkett and Rolls, 1998). John Carrol (2003) states that ‘an effective model of users judgement of information scent is a computational model based on spreading activation”, (Carrol, 2003, p. 174). Thus, a connectionist spreading activation model of language cognition was used to select keywords chosen to label the index cards. However, in these trials no formal computation was performed, but may be performed in the future. Fifty cards were produced in total. Once the cards were labelled with the keywords selected from the content analysis, some cards were kept blank and colored a shade distinct from the collection of fifty keyword cards. These colored cards were to be filled out by the subjects with category titles of their own choosing with a felt tipped marker. A voice recorder was used to record the sorting session while the subjects thought out loud about how best to arrange these cards into categories of information that was useful, and made sense. The category titles chosen by subjects were written on the coloured cards, (see Appendix 1 for an example of Card Sort Instructions). Analysis of this test included application of Vygotskian Activity Theory, and a consideration of the clustering of chosen cards between categories and across subjects.

The second test involved the use of a computer and internet browser, a video camera and tripod. Subjects were asked to complete three navigation search tasks on the CRaNHR site, (see Appendix 1 for instructions). Subjects were again asked to speak out loud while completing the tasks to provide an indication of any errors that occurred, and provide a sense of the rationale behind task completion. The tasks were modeled on business market specific Usability tests, for example, a subject in a market specific Usability test of a website owned by a shoe store may be asked to find the best priced pair of sandals. In this task, the subject would navigate the site until they were satisfied that they had found the best priced pair of sandals. The time that the subject took to complete the task and the errors that were made in reaching this goal would be noted and the recording would be analyzed to extract a rationale. In this study, the task goals were not sandals, but specific papers, reports and other scientific documents held at the CRaNHR website.

However, given that he CRaNHR site held so many research documents, we decided to provide some coordinates that would aid in the search, like a ‘geo-cache’ exercise where coordinates are given to players in order to find a buried treasure. Thus, the first two navigation tasks provided three possible locations where the goal could be found. The third and final question did not provide such coordinates, but left the entire website open to discovering the target item (See Appendix 1 for task instructions).

The final test was an administration of the System Usability Scale (SUS). This is a ten item questionnaire. This test was chosen because it reflects the ISO 9421-11 directives. The ISO 9421-11 states that the “effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments” should be considered (ISO 9421-11). The SUS was administered after the navigation task to provide a further measure of the navigation test.

Because most of the literature regarding HCI Usability is entrenched in market specific applications, the ISO guidelines, and much of the literature that treats HCI Usability Testing is evaluated according to tests that are significant to the marketplace. In order to strengthen external validity, methods common to market testing were used in evaluating data in this study. Measures of Efficiency were applied to the data according to the literature. Usability is defined as: “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” (Bevan, 2000). However, the idea that the scores from each of the Usability tests can be reduced to a single metric of Usability has been questioned by Saur and Kindlund (2005) who claim that “A single standardized and summated Usability metric (SUM) cannot be and should not the place of diagnostic qualitative Usability improvements typically found in formative evaluations.” (Sauro and Kindlund, 2005, p. 408). In this study the measures of “Task Completion”, “Effectiveness”, “Efficiency” and “Satisfaction” have been considered.

Procedure

The subjects were taken into a private room, asked to sit, and had the consent form read to them and thereafter asked to sign it. The first test was an Open Card Sort Test., This test is performed when you want to learn how users group content and understand the terms or labels users call each category. This test was explained to them and the instructions were provided on a separate sheet of paper for use during the test. The subjects were asked to sort the cards into groups that appeared ‘useful’ and ‘made sense’. This was performed on a wide desk top that allowed ample room for the spreading of the cards for analysis and grouping. The second instruction asked the subjects to name the categories that they had created by using a felt tipped marker to write the name of that category on coloured index cards. While the subjects were sorting the cards they were asked to ‘think out loud’ so that the decisions that they made in forming the groups would be recorded. The description provided during the task was recorded on a digital voice recorder, and provided insight into the rationale behind the task. When the subject was finished sorting the cards, they were asked to provide a brief summary of why they chose that particular name for each grouping. For example, if a group of cards were collected by a subject that dealt with universities and nursing programs, they may choose to call that category “Education”. (Please see Appendix 1 for an example of instructions). The fifty card each had a unique number on the back, and these numbers were entered into the cluster analysis. Similarities and differences between the groupings were extracted from the data. The similarities between the groupings will show how users commonly organized the terms, while differences showed where users had problems in understanding the keyword labels (Usability.gov; Card Sort).

After the Card Sort Test was completed, subjects were moved to be seated at a computer terminal. This is where the second test was performed. The second test was the Navigation Task and included use of a video camera that was focused on the screen to record the navigation process. Instructions were read to the subject. The instructions included a description of the objective of the test which was to evaluate the relations between various elements in the website.

It was described that the video camera was set up to record the navigation and especially the time taken for each task and the number and kind of errors produced in the navigation tasks. The task is to find the goal. There were three goals provided. Goal (A) was to find a paper titled “Primary Health Care Practitioners: Who are they? What do they do?”. It was decided that because the information on the CRaNHR site was very dense and poorly structured, that a ‘geo-cache’ approach would be adopted where coordinates, or ‘hints’ as to where the paper was located in the web site would be provided. In the first goal, three possible website locations were provided: the paper could be found under ‘research papers’, under ‘reports’ or under ‘focus documents’. The second goal (B) was similar to the first goal. However, the paper to be located in this case was “Sustainable Rural and Remote Primary Health Care Models in Australia”. The ‘hints’ to locating this paper included searching under “conference presentations”, “seminars” or “2008 news items”. The final goal (C) was to answer a question: “In what area does Michael J. Steedman perform his research?”. No ‘hints’ were given in this task. Subjects were expected to find the target without coordinates based on their experience from tasks (A) and (B). Subjects were permitted to ‘time out’ on a task at their own discretion if they could not locate this target. (Please see appendix 1 for instructions). The video tapes were then viewed and the navigation behavior was coded for analysis using Vygotskian Activity Theory (Kaptilinin, Nardi and MacAulay, 1999). The video tapes were scored on four Vygotskian parameters: Means/End (Hierarchical structure of activity), Environment (object – orientedness), Learning/cognition/articulation (externalization/internalization) and Development (development). Questions that guided the Activity Theory evaluation are found in Appendix 3. The video tape recordings provided a visual record of the search tasks as well as an audio record that included the subject’s descriptions of the rationale behind the search exercises. Both the visual record and the audio record were used in answering checklist questions. Moreover, the time that each subject took to complete each task was recorded. Moreover, task completion, task time and number of errors per task were applied in measures of Effectiveness and Efficiency based on task completion, error rates and time on task.

The final task was to fill out the SUS questionnaire. Subjects were removed from the computer position and sat at a desk to fill in the survey. The SUS survey consists of ten Likert scaled questions, (see Appendix 2). To calculate the SUS score, the scores from each item were summed. For items 2,4,6,8 and 10 the contribution is 5 minus the scale position. The sum of the scores were multiplied by 2.5 to obtain the overall system Usability score. This rating is also considered a satisfaction rating.

Measures of Effectiveness and Efficiency were taken from the scores on metrics used to analyze the site navigation of the CRaNHR website. Measures of Satisfaction were taken from the SUS survey. These metrics were considered together to produce a Productivity Measure that is in keeping with ISO measures of Usability that are in line with marketplace standards of measurement (Bevan, 2000).

Results

The data collected include results of three tests: a card sort test, a website navigation exercise and the SUS usability survey. The card sort test results were compiled in a connection table (Please see Table 1 for analysis results), that illustrated which of the fifty category cards were associated by the subjects, and which cards remained singletons. Table results indicate high connectivity between some content cards. The category “Laurentian University” was related to “Medical Education Rural” by a factor of three; “Long Term Care” and “Children/Adolescent Health” was associated by a factor of three; “PhD. Program in Rural and Northern Health” was associated with “Medical School Rural” by a factor of four; “Medical School/Education/Rural” was associated with “Continuing Medical Education” by a factor of three; and finally, “Medical School/Education Rural” was associated with “Lakehead University” by a factor of three, (See Appendices for Table). Singletons included “Knowledge Transfer”.

The website navigation exercise produced scores of “Success or Failure”, “Task Time” and “Number of Errors”. These scores were used to create measures of “Effectiveness: and “Efficiency”. A ceiling effect was noted across four of the subjects where at least one of the three tasks was not completed. Four of the measures across three subjects contained incomplete data due to malfunction of the recording equipment. Thus, a mean measure of “Efficiency” excluding malfunction effects across all subjects was 0.71 minutes per task as a measure of errors over completion time. In other words, of the tasks that were administered and completed successfully, the rate of completion while considering the number of errors produced is 0.71 minutes per (errors and completion), while the remaining tasks were unfinished. This measure is prescribed by the ISO criteria for website Usability Testing: “Efficacy relates the level of effectiveness achieved to the quantity of resources expended. Efficiency is generally assessed by the mean time taken to achieve the task.”, (Theofanos, M., 2007, H.12.2). In this present analysis, ‘resources expended’ is measured as ‘errors’ per completed task. However, Effectiveness alone is measured by the mean time that it took each subject to complete each task successfully. This mean number was 1.51 minutes/task. The lowest score, or the greatest number of errors per successful task completion time was 1.16 minutes/task. Of the 30 tasks administered, only 8 were completed successfully across all subjects while the rest failed due to the subject quitting the task or malfunction of recording devices, (Please see Table 2).

The System Usability Survey (SUS) produced a “Usability Metric” that included measures of subject satisfaction. The highest possible score on the SUS is 100. The mean score across eight subjects was 41.25. The highest score on the SUS was 82.5 while the lowest score was Zero (see Table 2), (Please see Appendix 2 for instrument and scoring information).

Table 1

	Adjacency – Card One	Adjacency – Card Two	Adjacency Factor	
Categories with High Adjacency in Cluster Analysis of Open Card Sort Test	Medical School/ Education Rural	Laurentian University	Three	
	Long Term Care	Children/Adolescent Health	Three	
	PhD. Program in Rural and Northern Health	Medical School/ Education Rural	Four	
	Medical School/Education Rural	Continuing Medical Education	Three	
	Medical School/ Education Rural	Lakehead University	Three	
	Medical Students Graduates	Medical School/ Education Rural		
Singleton				Knowledge Transfer

Table 2

Task	Effectiveness: Mean Task Completion Time per Task	Efficiency: Mean Task Time over number of Errors – Completion Rates	Satisfaction: Mean Post SUS score
Find this goal: It is a *research paper, a final report or a Focus Document*. “Primary Health Care Nurse Practitioners: Who are they? What do they do?”	6.3 Minutes	0.72 Minutes/(error completion)	
Find this goal: It is a *conference presentatation, a*	3.98 Minutes	0.96 Minutes/(error completion).	

seminar or a 2008 news item: “Sustainable Rural and Remote Primary Health Care Models in Australia”.			
Find this goal: In what area does *Michael J. Steedman* perform his research?	11.77 Minutes		
Mean Score			41.25

Discussion

The hypothesis that Usability Testing has provided data that may be used to improve the quality of Science Communication experience is supported. It is supported because subjects could not obtain adequate levels of “Satisfaction” in using the CRaNHR website for such reasons as inability to find desired documents and poor understanding of how to use the website in general. By identifying areas of weakness in Science Communication design, improvements may be made. Although there has not been very much work that investigates the relation between Science Communication and website Usability, there is a strong theoretical link between the two genres. Because authors in both fields recognize the importance and suitability of Vygotskian methods of analysis through Activity Theory, the premise that Usability Testing from the genre of HCI may be useful in producing data to treat Science Communication artifacts is tenable.

Experiment 1

The Open Card Sort Task provided subjects with an opportunity to organize website content headings in ways that they thought were useful and made sense. A cluster analysis of the data showed strong relationships between some content headings. The cluster analysis also showed the presence of singletons that were not related to any other content heading. Because some cards were strongly related by factors of three or four, those content headings may be seen to make the most sense and be most useful to subjects as they organized the categories. The presence of singletons indicates category headings that subjects found difficult to understand. Of interest, a singleton that was discovered in the analysis was “Knowledge Transfer”. This category card is also a Topic Heading used by the designers of the CRaNHR website. This is important to note because this means that the main topic heading used in the website was not understood by any of the subjects as it was not related to any other category card that was extracted from the website. Conversely, those cards that had a strong relation to other cards in the analysis show that subjects had a strong understanding of how to use that category title. The category cards that had the strongest relationship in the cluster analysis were the cards “Medical School/Education Rural” and “PhD Program in Rural and Northern Health”. These cards were related by a factor of four. Other cards that had a strong relationship were “Children/Adolescent Health” and “Long Term Care”; “Medical School/Education Rural” and “Continuing Medical Education”. These cards were related by a factor of two or three. All other cards except the singleton were related by a factor of two. Because these patterns were found in the Card Sort Test, it may be submitted that these patterns could be built from cognitive and social structures that are inherent across all of the subjects (Goforth, D. 2008). Because these structures can emerge in subjects to organize the scientific information inherent in the CRaNHR website, it can be submitted that the Science Communication theory specific to Informal Learning may be supported. Specifically, Vygotsky suggests that a “Zone of Proximal Development” may be attained to produce best learning results, especially in science learning activities. The “Zone of Proximal Development” is appropriate to science learning because it allows the ‘chunking’ of new information or conceptual strategies based on the presence of an acceptable ‘cognitive dissonance’ between previously learned information and new information, just as scientific knowledge is structured from accepted foundations to new ideas according to the hypothetico-deductive method. Because the analysis indicated that some cognitive and social structures may have emerged in the clustering, then one may submit that these structures are akin to the foundational structures that may be used to enhance Science Communication through Vygotskian methods, especially Constructivist applications (Hein, 2005). Moreover, because these structures emerge, it may be submitted that a Vygotskian social reconstruction of prior knowledge may have the ability to take place (Rochelle, 1995). Importantly, these strongly related categories suggest that the visitor may be able to quickly produce an ‘advanced organizer’ or cognitive strategy that may be applied to using the Science Communication system before the system has fully engaged. In this experiment, subjects were asked to use their own prior knowledge to sort through the semantic headings. A method of using ‘semantic networks’ to map associations among ideas has been employed in exercises of information processing and situated learning where these exercises have been applied to make it easier to describe prior knowledge precisely, (Rochelle, 1995, p. 46). The facilitation of such advance organizers through prior knowledge permits the construction of ‘Flow Experiences’ that are upheld by Czkszentmihalyi and Hermanson (1995) to be critical in producing motivation for informal learning and thus a successful Science Communication experience. Thus, the presence of strongly related category headings in the cluster analysis indicates that the relations between these category headings may emerge from cognitive and social structures proceeding from the subject’s prior knowledge, inducing successful informal learning experiences and stronger motivation to learn. Finally, this data has been produced by an experiment taken from the genre of HCI, but may be applied to informal learning and the improvement of Science Communication artifacts.

Conversely, the presence of a singleton in the cluster analysis indicates a poorly understood category heading. It is important to note that this category heading is a main searchable category taken directly from the CRaHNR website (Knowledge Transfer). Because there is no clustering, one can make the claim that there does not exist a conceptual structure within the subject sample that invokes a proclivity toward the website Usability or the quality of the Science

Communication experience. This negative finding is also important to the claim that the data from Usability Testing may serve to improve Science Communication experiences as some areas indicate need for improvement.

Experiment 2

The second experiment was a website navigation experiment that was recorded on video tape. There were three tasks in the experiment, each of which required the subject to find an object in the website, such as a research paper. Vygotskian Activity Theory was used to analyze the data and the rate of task completion, number of errors made in search tasks and the total search task time were also used in the analysis.

The Vygotskian Activity Theory analysis provided insights into how the subjects interacted with the CRaNHR website. Because the subjects were asked to 'think out loud', some indication of the rationale behind the experimental activities could be obtained. Points in the experiment where subjects changed their cognitive strategy in trying to find the target objects emerged from the Vygotskian analysis. One of the central tenants of Activity Theory is the identification of the production of Tools that are used by the subjects as they endeavour to transform their environment into a means of obtaining the target goal, (Kaptelinin, V., 1993). The identification of the strategy changes by the subjects provided these insights: Subjects would change strategies often in the course of the experiment by choosing to follow a different rationale in finding the target object. Moreover, subjects would often employ a computer 'hot key' that invoked a Graphic User Interface (GUI) designed to 'Find' a word or phrase in the text on the web page. These changes in strategy point toward the production of tools, or the use of such tools as the 'Find GUI' to help them meet the target goal. Failure of a strategy was counted as an error in the Data analysis. Of the 30 target goals administered in the experiment only eight were recorded as successfully completed. This means that the organization of the science information on the website did not succeed in producing a satisfactory Science Communication Experience or succeed in meeting respectable standards of HCI Usability. However, Activity Theory analysis indicates that as a response to strategy changes, subjects transformed the character of the CRaNHR website from a unified system where agglomerate information may be obtained (like a textbook), to a system where discrete information objects can be retrieved (like a library). The changes of strategies across subjects that lead to the formation of cognitive and social tools that encouraged search behaviors according to the 'library' model rather than the 'book' model produced shorter successful task times (See table 2). This means that the Vygotskian transformation of the HCI environmental system between the subject, the computer and apparent social ecology toward viewing the search goals as discrete items within the website rather than items whose contents were unified across the website produced a more successful Science Communication experience and a more successful User Interface. However, the overall low level of Success and low level of Effectiveness was further supported by vocalizations such as "This is impossible! I can't look through the whole website!", that were common across all subjects. Navigation tasks are inherent to system Usability, and it is submitted that the improvement of task times can be attributed to a transformation of the subject's environment and that these transformations are invoked through experience with the website producing 'good Vygotskian tool making'. Importantly, this finding supports the suggestion that the organization of the information within the CRaNHR website is not suited to successful Science Communication experiences in its current design form. Most importantly, the data from this HCI Usability metric has also shown that these kinds of analyses may be useful to the improvement of Science Communication artifacts.

Experiment 3

The last experiment in this study was the administration of the System Usability Survey (SUS). The survey was administered to all subjects and produces a 'Usability Metric' where a score of 100 would be the best Usability score. Because the mean score was 41.25, it is suggested that the Science Communication system within the CRaHNR website is less than half as good as other HCI that were used in developing the SUS. The SUS is immediately reflective of the ISO criteria for website Usability, so it may be claimed that the CRaNHR website is less than half as good as the international standard for website Usability. Note that the measures of the levels of Satisfaction are inherent in the survey, and the SUS score contains a viable measure of User Satisfaction, (See Table 2) (Please see Appendix 2). Because User Satisfaction is reflected in the SUS score, this Usability metric may also contribute to the improvement of Science Communication artifacts. For example, visitor satisfaction in Science Communication is inherently linked to successful learning outcomes and developing levels of motivation among visitors. Thus, a strong SUS score taken from a Science Communication artifact may indicate a high level of satisfaction.

Conclusion

All three experiments are linked by theory prevalent in Science Communication. Consideration of the ‘context’ that is inherent in the Science Communication experience has been analyzed within each experiment. This is because each experiment can be analyzed according to Vygotskian learning theory (Hein 2005), Rochelle’s (1995) ideas on prior knowledge and learning in interactive environments, Csikszentmihalyi and Henderson’s (1995) notions of intrinsic motivation and Hedge’s (1995) ideas on Human-Factors considerations on museum Science Communication artifacts. All of these authors consider ‘context’ in a different way, from a Vygotskian social context to the influence of ‘context’ in Csikszentmihalyi and Henderson’s (1995) ‘Flow Experiences’. Moreover, literature within the genre of HCI evaluations include disclosure on the use and application of Vygotskian Activity Theory in performing Usability analyses. Thus, it is clear that Vygotsky is a pivotal link between the theory that is applied to both genres. This necessarily gives rise to the applicability of HCI Usability Testing in evaluating Science Communication artifacts. However, the HCI Usability Testing is not subject to the same kind of treatment for external validity as the research in Science Communication. The research in Science Communication is subject to the accepted norms of peer evaluation and repeatability to provide a basis for external validity. However, the HCI literature follows an international standard that has been developed by professionals within the marketplace. For this reason, testing design is heavily selected by economic factors rather than scientific foundations of external validity. The market driven character of the HCI evaluations have selected for such measures as Efficiency, Effectiveness, and Satisfaction as key determinants of Usability and are intrinsic to the International Standards Organization criteria for website Usability (ISO 9421). Thus, the results of this study are supported by such authors in the HCI genre as Neilson (2004), Schneiderman (2005), and Bevan (2000).

The Center for Rural and Northern Health Research website is a Science Communication artifact because the product that CRaNHR produces is scientific experimentation, literature, reports and studies. This study concludes that by using HCI Usability Testing the quality of the Science Communication can be rigorously assessed. Moreover, that assessment may imply that the quality of the Science Communication experience for the CRaNHR artifact is below international industry standards for website Usability. Moreover, Science Communication theorists such as Rochelle (1995) and Cszicksentmihalyi and Henderson (1995) have provided literature that supports this study's claim that one may use HCI Usability Testing as a viable metric for Science Communication because these tests measure what these authors describe. For example, because of poor information design evidenced by the Usability Testing, the CRaNHR website cannot maximize the use of such essential informal learning components as 'prior knowledge' and 'motivation', which are examples of Rochelle's (1995) and Cziksentmihalyi and Henderson's (1995) theories. This study does not provide solutions to the CRaNHR design problem, but has helped to solve the problem of how to rigorously measure Science

Communication in the web site medium. Moreover, this study has provided new conclusions and theoretical implications such as the significance of Science Communication theory in evaluating category clustering in Science Communication networks such as web site networks. The formation of 'Vygotskian Tools' is similar to the production of Goforth's (2008) concept of an 'epitome' in interactive Science Communication networks and artifacts. This is especially interesting in the case of examining high levels of adjacency in cluster analyses. This is because this study has provided some data that supports the use of cluster analyses of information systems in order to extract an 'epitome' using standard HCI Usability Testing such as Open Card Sort analysis. In this way the relation between experiments 1 and 2 may be illustrated because the Open Card Sort Task, Navigation Task and Vygotskian analyses support the emergence of an 'epitome'. Of importance, the application of market selected Usability measures such as Effectiveness, Efficiency and Satisfaction to the use of such a tool, or epitome, is new and rigorous way to evaluate the tenor of such user centered construct. This means that if such an epitome were to be identified, the third Experiment would assess the epitome according to market measures, such as improvements in Efficiency, Effectiveness, and Satisfaction. Furthermore, the Vygotskian notion of transformation of the ecology of the User-Computer Interface, or the Visitor-Science Communication artifact engagement, is also elucidated by the Usability Testing. Finally, the importance of examining Science Communication in information networks is growing: Management of scientific information within sophisticated networks is becoming more pervasive and the previous methods of providing Science Communication experiences in linear information structures (like learning from a book) is less common. Treating Science Communication evaluations with methods that succeed in network environments, such as HCI Usability Testing is essential to meet the upcoming demands of our culture and is very suitable to Constructivist learning scenarios such as those found in science museums. However, this study as only provided an initial examination of this topic and much more research must be completed to fully develop all the opportunities that the use of HCI Usability Testing may provide to the improvement of Science Communication.

Bibliography

American Psychological Association (1994). Publication Manual of the American Psychological Association. Fourth Edition., Washington D.C.: American Psychological Association.

Bell, Judith., (2005). Doing Your Research Project. Fourth Edition. New York, NY: Open University Press.

Bevan, Nigel (2000). *ISO and Industry Standards for User Centered Design*. Retrieved May Second, 2008, from http://www.idemplyee.id.tue.nl/g.w.m.rauterberg/lecturenotes/OH420/Usability_standards .pdf

Bolter, Jay David, and Gromla, Diane., (2003). Windows and Mirrors. Cambridge Massachusetts, Massachusetts Institute of Technology.

Carroll, John M., (Ed.)., HCI Models, Theories and Frameworks. Toward a Multidisciplinary Science. San Francisco, CA: Morgan Kaufmann Publishers.

Creswell, John W., (1994). Research Design. Qualitative and Quantitative Approaches. Thousand Oaks, CA: Sage Publications Inc.

Csziksentmihalyi, Mihaly and Hermanson, Kim., (1995). Intrinsic Motivation in Museums: Why Does One Want to Learn?. In Falk, John H. and Dierking Lynn D., (Eds.) Public Institutions for Personal Learning., (pp. 67-77). Washington D.C.: American Association of Museums, Technical Information Service.

Contextual Interviews, Retrieved June 27, 2008
http://www.usability .gov/design/prototyping.html

Dede, Chris, Ketlhut, Diane, Ruess, Kevin (No year given). *Designing for Motivation and Usability in a Museum-Based Multi-User Environment*. Retrieved June 7th, 2008, from http://muve.gse.harvard.edu/rivercityprojectg/documents/AELppr.pdf

Develop a Prototype. Retrieved June 27th, 2008
http://www.usability.gov/design/prototyping.html

Dumas, Joseph S., Reddish, Janice C., (1999)., A Practical Guide to Usability Testing., Portland, OR: Intellect Books.

Eden, J., (No Date Given). *Agent Based Modeling of Usability from a Distributed Cognitive Perspective.* Retrieved 6\28\2008 from http://agent2007.anl.gov/2007pdf/Paper%2030%-%2Eden_Agent2007_Final.pdf

Focus Groups. Retrieved June 27th, 2008 http://www.usability .gov/methods/focusgroup.html

System Usability Score Applied to ENRICH Project Web Site. Retrieved June 28, 2008 from http://musicalia.IT/sus/results.htm

Faulkner, Christine, .(1998). The Essence of Human-Computer Interaction., Heel Hempstead Herefordshire, UK: Prentice Hall Europe.

Goforth, David, (2008). Class notes and personal communication.

Harms, Ilse, Schwiebenz, Werner. (2001). *Evaluating the Usability of a Museum Website.* Retrieved June 7th, 2008 from http:scidok.sulb.uni-saarland.de/voltexte/2007/864/index.html

Hein, George, E., (1998). Learning In the Museum. New York, NY: Rutledge.

Hedge, Alan J., (1999). Human-Factor Considerations in the Design of Museums to Optimize Their Impact on Learning. In Falk, John H., and Dierking, Lynn D., (Eds.) Public Institutions for Personal Learning., (pp 67-77). Washington D.C.: American Association of Museums, Technical Information Service.

Helfrich, Alan J., (1999). Building On Ramps to the Information Super Highway: Designing, Implementing, and Using Local Museum Infrastructure. In Schiele, Bernard and Koster, Emlyn, H., (Ed.), Science Centers for This Century., (pp. 87-123). Sainte-Foy PQ: Editions MultiMondes.

Honeyman, B., (2001). Real vs. Virtual Visits: Issues for Science Centers. In Errington, S., Stockelmayer, S.M., Honeyman, Benton, (Eds.) Using Museums to Popularize Science and Technology. (pp. 107-11). Pall Mall, UK: Commonwealth Secretariat.

Heuristic Evaluation. Retrieved June 27th, 2008 http://www.usability.gov/methods/heuristic eval.html

Huang, Weihong., , Webster, David., Wood, Dawn., Ishaya, Tanko., (2006). *An Intelligent Semantic E-Learning Framework Using Context Aware Semantic Web Technologies.* Retrieved July 7., 2008 from http://cat. Inist.fr/?aModele=afficeN&cpsidt=17713762

Individual Interviews. Retrieved June 27th, 2008 from http://www.usability.gov/methods/individual .html

Kaptelinin, Victor, (1993). *Activity Theory: Implications for Human Computer Interaction.* Retrieved July 25, 2008 from
http://www.comp.leeds.ac.uk/umas/reading-group/kaptelinin-ch5.pdf

Kaptelinin, Victor, Nardi, Bonnie, MacAulay, C., (1999). *The Activity Checklist: A Tool for Representing the "Space" of Context.* Retrieved July 28, 2008 from
http://delicery.acm.org/

Krippendorff, Klaus., (2004). Content Analysis. An Introduction to its Methodology. Thousand Oaks CA: Sage Publications Inc.,

McLeod, Peter., Plunkett, Kim., Rolls, Edmund T., (1998). Introduction to Connectionist Modeling of Cognitive Processes. Oxford, UK: Oxford University Press.

Nardi, Bonnie, A., (1998). *Activity Theory and its Use Within Human Computer Interaction.* Retrieved July 25, 2008
http://www.jstor.org/stable/1466817?seq=5

Neilson, Jakob and Loranger, Hoa., (2006). Prioritizing Web Usability., Berkeley, CA: New Riders.

Neilson, Jacob. (2004). *Information Foraging. Why Google Makes People Leave Your Site Faster.* Retrieved June 22, 2008 from http://www.useit .com/alertbox/20030630.html

Perform Card Sorting. Retrieved June 27, 2008
http://www.usability.gov/design/cardsort.html

Plano, Jennifer, Rogers, Yvonne, Sharp, Helen., (2002). Interaction Design. Beyond Human-Computer Interaction. New York, NY: John Wiley & Sons Inc.

Ratner, Julie. (2003). Human Factors and Web Development, Second Edition., Mahwah NJ: Lawrence Erlbaum Associates Inc.

Rochelle, Jeremy, (1995). Intrinsic Motivation in Museums: Why Does One Want To Learn?. Learning in Interactive Environments: Prior Knowledge. In Falk, John H., and Dierking, Lynn D., (Eds.) Public Institutions for Personal Learning. (pp. 67-77). Washington D.C. : American Association of Museums, Technical Information Service.

Russel, David R., (2001). *Looking Beyond the Interface: Activity Theory and Distributed Learning*. Retrieved August 12, 2008 from
http://72.14.205.104

Rout, Jean-Francois, Levonen, Jarmo, J., Dillon, Andrew., and Spiro, Rand J., (Eds.) (1996). Hypertext and Cognition. Mahwah, NJ: Lawrence Erbaum Associates, Inc.

Stanard, Terry, Hutton, Robert, J.B., Warwick, Walter, McIllwaine, Stacy, McDrmott, Patricia L., (No Date). *A Computational Model of Driver Decision Making at an Intersection Controlled by a Traffic Light*.., Retrieved June 27th, 2008 from http://ppc.uiowa.edu/driving-assessment/2001/Summaries/Driving%20Assesment%20Papers/65_stand.pdf

Theafanos, M., (Ed.) (2007). *Common Industry Specifications for Usability Requirements.* *http://zing.ncsl.gov/design/paralleldesign.html*

Use Parallel Design. Retrieved June 27th, 2008 from http://www.usability.gov/iusr/documents/CISU-R-IR7432.pdf

Quek, Amanda, (No Date). *Toward an Activity Theoretical Evaluation Method for Web Based Systems.* Retrieved July 25, 2008 from http://www.comp.lancs.ac.uk/computing/research/cseg/projects/tracker/quek03.pdf

Appendix 1

Task Instructions for Participants

CRaNHR Website and Science Communication

Task 1: Card Sort

We are trying to determine what categories of information will be ***useful***, what groupings make ***sense***, and what the groupings should be ***called***.

1. Sort the cards and talk out loud while doing so. Talk about the reasons why you have chosen to place that card in the group. We are looking to find the ***rationale*** behind the sorting.
   Additional content cards may be named and added as you think necessary during the sorting process.

2. ***Name*** each grouping on the colored blank cards using words that you would expect to lead to that particular grouping.

Task 2: Timed Navigation Task

We are trying to identify relationships that exist between various elements in the website. By videotaping the computer screen, we will be recording to measure the ***time*** that it takes you to complete a task and the number of ***errors*** you make (dead ends or back-tracking). Your task is to find the ***goal***.

1. Navigate toward the goal and talk out loud while doing so. Explain why you are choosing that route, mention if you have reached a dead end or have made some error in navigating until you reach the ***goal***.

2. To begin, turn on the monitor, this will start the clock. To stop the clock when you have completed the task, turn off the monitor.

Task Goals

a. Find this **goal**: It is a *research paper*, *a report or a Focus* document.

   “Primary Health Care Nurse Practitioners: Who are they? What do they do?”

b. Find this **goal:** It is a *conference presentation*, *a seminar or a 2008 news item:*\

“Sustainable Rural and Remote Primary Health Care Models in Australia.”

c. Find this **goal**:

In what area does *Micheal J. Steedman* perform his research?

Appendix 2

How to Score the SUS

To calculate the SUS score, first sum the score contributions from each item. For items 1,3,5,7, and 9 the score contribution is the scale position minus 1. For items 2, 4, 6,8 and 10, the contribution is 5 minus the scale position. Then each item’s score contribution will range from 0 to 4. If we multiply the sum of scores by 2.5 the overall value of the SUS obtained will have a range from 0 to 4. If we multiply the sum of the scores by 2.5 the overall value of the SUS obtained will have a range of 0 to 100.
SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own.