Supporting Elementary-Age Children's Searching and Browsing: Design and Evaluation Using the International Children's Digital Library

Hilary Browne Hutchinson[1,*], Allison Druin[1,2], Benjamin B. Bederson[1]

Human-Computer Interaction Lab

Department of Computer Science[1] and College of Information Studies[2]

University of Maryland, College Park, MD 20742 USA

{hilary, bederson}@cs.umd.edu, allisond@umiacs.umd.edu

* First author current contact information:

Google, 1600 Amphitheatre Parkway, Mountain View, CA 94043

ABSTRACT

Elementary-age children (ages 6-11) are among the largest user groups of computers and the Internet. Therefore, it is important to design searching and browsing tools that support them. However, many interfaces for children do not consider their skills and preferences. Children are capable of creating Boolean queries using category browsers, but have difficulty with the hierarchies used in many category browsing interfaces because different branches of the hierarchy must be navigated sequentially and top-level categories are often too abstract for them to understand. Based on previous research, we believed using a flat category structure, where only leaf-level categories are available and can be viewed simultaneously, might better support children. However, this design introduces many more items on the screen and the need for paging or scrolling, all potential usability problems. To evaluate these tradeoffs, we conducted two studies with children searching and browsing using two types of category browsers in the International Children's Digital Library. Our results suggest that a flat, simultaneous interface provides advantages over a hierarchical, sequential interface for children in both Boolean searching and casual browsing. These results add to our understanding of children's searching and browsing skills and preferences and also suggest guidelines for other children's interface designers.

MOTIVATION

In both the European Union and the U.S., households with children are more likely to have computers and Internet access than households without children (Day et al., 2005; Demunter, 2005). In 2003, 42% of U.S. children age 5-9 and 67% of children age 10-13 used the Internet (NTIA, 2004). Around the world, children use the Internet for schoolwork, to play games, and to communicate with each other (Day et al., 2005; Livingstone, 2003). These activities often require searching and browsing, similar activities distinguished by a methodical approach with a specific goal in the former and an emphasis on progressive filtering of results based on visual scanning in the latter (Ahlberg & Shneiderman, 1999). Search engines and digital libraries with keyword or category-based interfaces are two of the most common, freely-accessible tools used to support both of these activities on the Internet. Search engine sites such as Yahooligans! (*yahooligans.yahoo.com*) and Ask Jeeves Kids (*www.ajkids.com*) are examples of portals that children can use to find age-appropriate content. Project Gutenberg (*www.promo.net/pg*) and the Rosetta Project (*www.childrensbooksonline.org*) are examples of large digital libraries with books for children. However, these and other web sites for children often have interfaces with one or more of three problems.

First, they do not take into account the information processing and motor skills of children, specifically their difficulties manipulating small objects with a mouse (Strommen, 1994; Inkpen, 2001; Hourcade et al., 2003). Second, they do not consider children's searching and browsing skills, specifically their difficulties with spelling, typing, and composing keyword queries (Moore & St. George, 1991; Borgman et al., 1995; Druin, 2005); understanding and navigating category structures (Rosch et al., 1976; Tversky, 1985; Marchionini & Teague, 1987); and understanding and creating Boolean queries (Neimark & Slotnick, 1970; Tversky & Kahneman, 1975; Reuter & Druin, 2005). Third, they do not consider how children prefer to search, presenting search criteria appropriate for adults, but not for children, who often like to search differently (Wendelin & Zinck, 1983; Moore & St. George, 1991; Kragler & Nolley, 1996; Druin, 2005).

BACKGROUND

Recent work in our lab has focused on designing category browsers for digital libraries that support elementary-age children's abilities to search and browse. The QueryKids interface allowed children to find information about animals in a zoomable user interface (Druin et al., 2001). This interface included large, easily clickable icons rather than a keyword search box that required typing; incremental and clearly visible results to show search progress; and a built-in Boolean protocol to prevent children from having to mentally construct Boolean queries. Search categories were based on how children liked to search for animals, such as what they ate or where they lived. Revelle et al. (2002) found that second and third grade children (ages 7-9) were able to use this interface to conduct both simple and Boolean searches 85% of the time.

Based on the success of the QueryKids interface, we built the International Children's Digital Library (ICDL, *www.icdlbooks.org*) (Druin et al., 2003; Druin, 2005), a free collection of over 1000 children's books from around the world used each month by over 70,000 people from more than 170 countries. The first version of the ICDL used an interface similar to QueryKids, with the addition of a hierarchical category browser to find books (Figure 1).

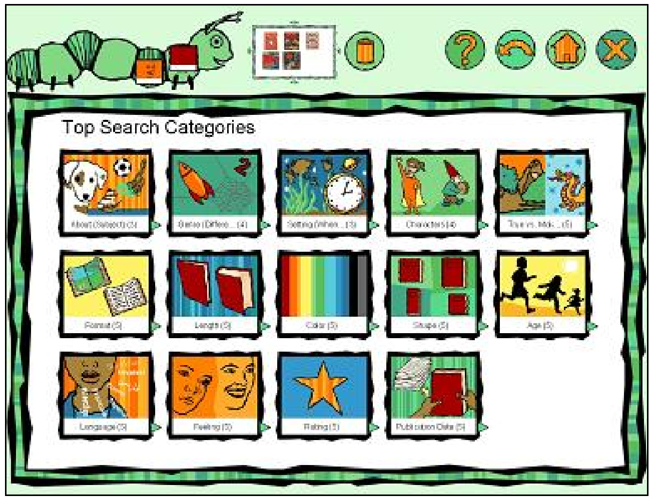

Figure 1. Original ICDL category browser

The original ICDL category browser was created using *faceted metadata* (Yee et al., 2003), a collection of independent classifiers such as shape, color, language, and genre, each of which was hierarchical in structure. The 14 top-level facets were presented using *sequential menus*, in which only one facet at a time could be explored. Users could navigate down into a facet to select leaves for a search. Creating a Boolean search using categories from different facets required navigating to the leaves of one facet to make a selection, backtracking to the top of the hierarchy, and then navigating to the leaves of another facet. Leaves with the same parent were combined disjunctively (OR), while leaves with different parents were combined conjunctively (AND). This implicitly supported Boolean queries such as "(Red OR Orange) AND English". Selected leaves were moved to a caterpillar "helper" at the top of the screen to keep track of the search. Books that matched the search were shown in a small box next to the caterpillar that could be zoomed in for closer inspection once the search was complete.

Reuter & Druin (2005) found that children in grades one through five were able to navigate the category structure to find books in open-ended browsing, but they did not generally use the Boolean capability on their own. Based on this study, we believed that the structure and presentation of the category browser discouraged children from creating Boolean queries. While Boolean search is known to be difficult for both children and adults, children are capable of using it (Neimark & Slotnick, 1970; Tversky & Kahneman, 1975). As the size of the library grew, it became more important to better support Boolean queries, especially conjunctive Boolean queries, to help users narrow down their results.

Therefore, we redesigned the category browser to see if we could better support children creating Boolean queries using the ICDL (see Hutchinson et al., 2006 for more design details). In the paper that follows, we discuss related work, describe the changes made to redesign the browser, present the results of two studies we conducted to evaluate these changes, and suggest design guidelines for others creating searching and browsing interfaces for elementary-age children.

RELATED WORK

The interface design decisions for the ICDL category browser were based on related work in three areas: information processing and motor skills; searching and browsing skills; and selection criteria. First, young children process information more slowly than adults, and this affects their motor skills (Thomas, 1980; Kail, 1991). For motor skills that involve a mouse, early elementary age children are slower to acquire a target than adults and require larger targets (Hourcade et al., 2003). Early elementary age children also have difficulty holding down a mouse button for extended periods of time, coordinating dragging and clicking, double-clicking, and using multi-button mice (Strommen, 1994; Inkpen, 2001; Hourcade et al., 2003).

Second, children have difficulty with the two most common searching and browsing tools: keywords and category hierarchies. Keyword queries are problematic for children at all elementary school ages because they have difficulty spelling, typing, and composing queries (Moore & St. George, 1991;

Borgman et al., 1995; Druin, 2005). Category hierarchies are problematic for several reasons. Mid and late elementary age children do not always navigate them efficiently (Marchionini & Teague, 1987). Pre-school and early elementary age children may not think hierarchically yet, preferring to organize items perceptually (Tversky, 1985). Early elementary age children also have difficulty understanding abstract, top-level categories (Rosch et al., 1976). Both children and adults also have difficulty creating Boolean queries with keywords (Borgman et al., 1995). Mid elementary age children are capable of completing Boolean searches in hierarchical category browsers (Revelle et al., 2002), but do not seem to create Boolean queries spontaneously when browsing (Reuter & Druin, 2005).

Finally, children look for books differently than adults. In physical libraries, early elementary age children choose books based on the appearance of the cover and illustrations (Moore & St. George, 1991; Kragler & Nolley, 1996; Druin, 2005). Older children focus on textual summary information in book jackets, covers, and indices (Wendelin & Zinck, 1983). Early elementary age children tend not to make a distinction between fiction and non-fiction books, and prefer books about certain genres like fantasy over fiction or learning books (Kuhlthau, 1988). Older children focus on particular genres that interest them, such as sports and animals (Wendelin & Zinck, 1983; Kuhlthau, 1988). Recommendations by peers and teachers also have an important influence on children’s book selections (Kragler & Nolley, 1996). Many of these trends are also true in digital libraries (Druin, 2005; Reuter & Druin, 2005).

Several digital library systems for children have attempted to address these issues (e.g. Pejtersen, 1989; Busey & Doerr, 1993; Borgman et al., 1995). These systems included large, easily clickable buttons and child-appropriate categories, addressing the first and third concerns described above. However, they were generally based around a hierarchical category browser using the Dewey Decimal system, an improvement over keyword-based systems, but still a concern because of the hierarchy. Project Gutenberg and the Rosetta Project are current examples of public digital libraries with children's books, but these have adult-oriented interfaces. A number of digital library catalogs for children with more appropriate interfaces exist (e.g. Kids Catalog Web, Storyplace), but don't actually provide access to books. The original ICDL category browser also addressed the first and third concerns, and provided full access to books. However, while it had a hierarchy with more child-appropriate terms than the Dewey system, it still suffered from the other challenges associated with hierarchies.

NEW INTERFACE

We believed the problems with the original ICDL category browser could be fixed with changes to both the structure and presentation of the interface. For structure, by *flattening* the depth of the facet hierarchies, we believed that it would make it easier for children to navigate, as has been found for adults (Miller, 1981). We believed the benefit would be even greater for elementary-age children, who, depending on their age, are still learning hierarchical categorization, don't always navigate hierarchies efficiently, and may have trouble understanding abstract, top-level categories. For presentation, we believed that using a *simultaneous* presentation of categories, in which each facet or branch in a category structure can be explored in parallel, might be faster and easier for children to use to create Boolean searches. For adults, it was shown that a simultaneous browser could be used to facilitate faster creation of Boolean searches than a sequential browser (Hochheiser & Shneiderman, 2000).

Based on these hypotheses, we designed a new interface, which consisted of a ring of category icons arranged around a collection of books (Figure 2). First, we flattened the hierarchies in the original category facets to a single layer, using the leaf-level categories rather than the abstract, top-level categories. Next, we analyzed a year's worth of web log data and identified a subset of the most popular of these categories and presented them simultaneously, avoiding much of the navigation required in the sequential presentation. Finally, we placed the search results in the middle of the screen to draw more attention to the goal of the search: reading the books in the library.

Figure 2. New ICDL category browser

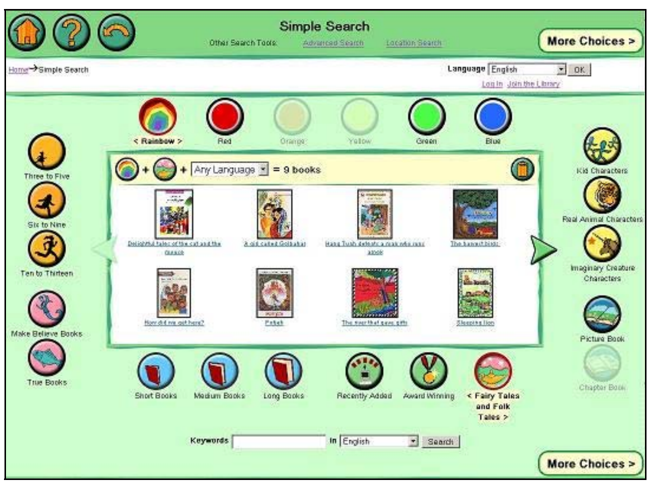

Figure 3. **New interface after selecting Rainbow and Fairy Tales categories**

Because the category icons needed to be large enough for children to click on, we could not fit all the categories on a single screen. Hourcade et al. (2004) found that icons with 64-pixel diameters are sufficient for children as young as 4 to be able to click reliably, so we chose icons of this diameter for categories. As a result, we had to introduce either paging or scrolling to accommodate them. We chose paging (between 2 pages) because research has shown it to be superior to scrolling in many situations (Mills & Weldon, 1987), and because we wanted the interface to fit on a single screen at a common 1024x768 pixel resolution. This resulted in a design with 22 categories on the first page, plus two buttons labeled “More Choices” that linked to a second page with another 22 categories. More Choices buttons in the same locations on the second page link back to the first page.

We followed the advice of Plaisant et al. (1997) to bring the “treasures” of the library to the surface by having books that match the selected categories appear prominently on the same page with the search tools, rather than in a small corner or on another page, as we did in the old version. When no categories are selected, a group of 2 or 3 featured books appears in the results area. We placed the books prominently in the center of the page because we felt that they should be the main focus of the page. If a child doesn’t understand how to use the category browser, he or she can still find some books to read. The inspiration for this design, which we originally called Fisher-Price®, came from our observation of toys for young children, which often have a central feature with large buttons around the outside. We felt that using this familiar design might make children more comfortable.

Clicking a category button selects it and adds it to the search. Clicking it again unselects it and removes it from the search. The results are incrementally updated whenever new categories are added or removed. For instance, if a user selects the *Rainbow* and *Fairy Tales* categories, the results show books that match both categories (Figure 3). Categories that do not match any books in the current result set are grayed out to avoid no-hit search results. When categories are selected, they are combined in an “equation” across the top of the results section to indicate that their combination adds up to the count of the results. This visual tool makes the effect of selecting multiple categories concrete, which is important for children learning to reason logically (Shneider, 1996). In addition to selecting categories, users can refine their search by including keywords and limiting the results to books written in a particular language. The keyword appears as part of the search equation. The language menu also appears in the equation and contains only languages that appear in the current result set, preventing the creation of no-hit searches. However, it is possible to create no-hit searches when keywords are included in a search.

The selected categories, keywords, and language are all joined conjunctively. We chose to support only conjunctive Boolean searches for three reasons. First, children have an easier time with conjunction than with disjunction (Neimark & Slotnick, 1970). Second, the goal of the interface is to narrow down the number of books from a large collection so that children can easily select from a few. Conjunction will decrease or keep constant the number of results. Finally, while the original category browser used conjunction between categories and disjunction within category groups, we felt that this would be confusing when all the categories are presented together on the same level.

The resulting interface addressed the concerns we had with the original interface. Children can rely on perception rather than hierarchical knowledge to find categories because the hierarchy is flattened. Children need only select from concrete, leaf-level categories because the more abstract, top-level categories are removed. They do not have to navigate and backtrack in a hierarchy because the categories are viewed simultaneously. However, this design had three important tradeoffs. First, the increased number of categories on the page could be visually overwhelming to young children (Shneider, 1996). We nearly doubled the number of categories that appear on the page from the original browser. Second, placing the categories around the perimeter resulted in an interface that requires a lot of visual scanning to see them all. More traditional interfaces of this type place categories on one side and results on the other side to reduce visual scanning. Finally, by using paging to accommodate all the categories, we ran the risk of children not finding the paging buttons and never reaching the second page of categories.

Despite these reservations, we still believed that this design would be superior, and we replaced the original browser with this new design in October 2004. To evaluate our decision, we formally explored the tradeoffs for children completing searching and browsing tasks using a hierarchical structure presented sequentially, similar to the old ICDL category browser, and a flat structure presented simultaneously, like this new version, with two studies. Until now, no studies have looked systematically at how children of different ages are able to use hierarchical and flat category structures, simultaneous and sequential presentations, and Boolean logic.

EVALUATION

We conducted two studies to evaluate the changes we made to the structure and presentation of the new ICDL category browser. We created two interface designs to compare these changes and had children in three different age groups evaluate them. The first design, *flat,* was similar to the new ICDL category browser with a single level of categories and required paging to reach half of the categories. The second design, *hierarchical,* was similar to the old ICDL category browser and had a two-level hierarchy that required sequential navigation to reach categories in different facets. We were mainly interested in whether the flat design supported creation and understanding of Boolean queries better, but also wanted to understand whether it was any better for simple, single category queries. Because we varied both the structure (flat vs. hierarchical) and presentation (simultaneous vs. sequential) in the two designs, we were not able to study which of these changes, if any, might have had a larger effect.

In both studies, children first completed loosely timed, free browsing tasks where they could select as many categories as they wanted. In the first study, children were then asked to do searches with each interface involving a single category (*simple study*). In the second study, children were asked to do searches with each interface involving two categories (*Boolean study*), creating conjunctive Boolean queries. Children were timed on these searching tasks and we recorded the number of hints they required to successfully complete each task. Finally, children were asked questions to elicit information about comprehension and perceived likeability, difficulty and preference. Both studies followed essentially the same protocol, described below.

Hypotheses

We had the following hypotheses for our studies:

H1. The flat interface would better support creation of multiple-category (Boolean) queries while browsing than the hierarchical interface.

H2. The flat interface would be faster than the hierarchical interface for simple searches that did not require paging in the flat interface.

H3. The flat interface would be faster and require fewer hints than the hierarchical interface for Boolean searches.

H4. The flat interface would be easier, more likable, and preferred to the hierarchical interface for Boolean searches.

H5. Boolean searches would be harder for younger children to complete than for older children.

H6. The flat interface would better support understanding of Boolean logic in Boolean searches than the hierarchical interface.

Study Designs

Participants

The participants were 72 first grade (age 6-7), third grade (age 8-9) and fifth grade (age 10-11) children. In each study, 36 children participated, 12 in each grade, equally split between boys and girls. Children came from after-school programs and regularly scheduled technology classes in 4 suburban Maryland elementary schools. All were familiar with computers; none had previously used the ICDL.

Materials

Participants used a Dell laptop with a 12 inch display, 1024x768 screen resolution, an Intel Pentium 3 Mobile CPU running at 800 MHz, Windows XP, and a Kensington single-button mouse. The software was an adapted version of the ICDL running locally and instrumented to record the time of each mouse click. Participants used the Microsoft Internet Explorer 6 browser in full screen mode with the task bars hidden.

Interfaces

Both interfaces allowed children to search for books using category buttons that could be selected and unselected to include or exclude them from the search. The keyword and language features from the new interface design were not included. Categories were arranged around the perimeter of the screen and books were presented in the middle in sets of 8. The *flat interface* used a single-level category structure (Figures 4, 5). It presented 44 leaf-level categories over two pages. It required paging by use of “More Choices” arrows at the top and bottom of the screen. The *hierarchical interface* used a two-level category structure (Figures 6, 7). It organized the same 44 leaves into 9 top-level facets (Color, Length, Format, Age, Characters, True vs. Make Believe, Genre, Rating, and Feeling), each with between 2 and 12 leaves on the second level. To select leaves from different facets, users had to backtrack using a large

"Up Arrow". In both interfaces, 3 featured books were shown in the results area if no categories were selected.

Figure 4. Page 1 of the flat interface

Figure 5. Page 2 of the flat interface

Figure 6. The top level of the hierarchical interface

Figure 7. The color leaf level of the hierarchical interface

Procedure

In each study, participants worked with a researcher one at a time in a quiet room or hallway in the school. Participants completed browsing, searching, comprehension, difficulty and likeability tasks with one interface, then the other. Each participant used both interfaces, so we created two different but structurally and cognitively similar sets of searching tasks for each study. For example, children might be asked to find Red books in one set and Orange books in the other. These category buttons were located near each other in both interfaces and involved the same concept. Interface order and task set use were counterbalanced and the order of the searching tasks in each task set was randomized for each child.

Tasks

Browsing Task

In both studies, the first task was a loosely timed, free browsing task to observe how participants used an interface without instruction. Children were asked to try out the program to find some books. After the children explored for two or three minutes, the researcher took control of the mouse and demonstrated the interface. This included scripted instructions on how to select a single category, how to unselect categories and start over, how to use the results paging arrows, and how to use the More Choices or Up

Arrow buttons. It did not include instruction on how to create a Boolean query, since one of our study goals was to observe whether children understood this feature.

Searching Tasks

Participants then completed six timed, randomized search tasks. Each search task was a question of the form "How many X books are there?", where X was one category in the simple study and two categories from different facets in the Boolean study. In the flat interface, half of these tasks required using the second page (*type A tasks*) and half did not (*type B tasks*). In the hierarchical interface, there was no functional difference in completing these two types of tasks, but they are labeled as such so that they can be compared with the same tasks in the flat interface.

Comprehension Tasks

In the Boolean study, children completed two additional tasks in which they were asked to select two category buttons and then questioned about what kinds of books were found. One task was completed before the six search tasks and one was done immediately after. These two questions were designed to elicit whether participants understood that they were creating a conjunctive Boolean query when they selected two category buttons. A strict protocol for questioning children was developed to elicit these responses.

Subjective Tasks

After completing the tasks in an interface, participants answered two subjective questions about the difficulty and likeability of the interface. Participants were asked if the interface was hard, medium, or easy to use, and whether they liked using it not much, a little, or a lot. Finally, after participants had used both interfaces, they were asked if they liked one interface better than the other, or both the same.

Hints

University rules required that the study take no more than 30 minutes per child, which pilot testing showed would require giving some children hints until they successfully completed a task, or enforcing time limits during search tasks. We chose to give hints as verbal scaffolding (Soloway et al., 1991) to support children so that we would have complete time data. We have observed that working with an adult is a common use scenario for elementary-age children using the ICDL, and 76% of children age 6 to 12 report that there is an adult in the room while they are online (CPB, 2002). Children also often work together on both computer and non-computer projects in school environments, and often receive help from their peers in this way (Druin et al., 1997; Stewart et al., 1999; Clements et al., 1999).

During pilot testing, common problems with the search tasks were identified and strict protocols were developed for giving hints for each problem to reduce the potential for bias. For instance, if a child selected the wrong category in a search task, the researcher went through four gradually more helpful hints until the child was able to identify and correct the problem. For the first hint, the researcher indicated that they didn’t have the right answer, and asked if they saw why. If not, the researcher pointed out which category they had selected and reminded them of which category they were actually looking for. Finally, if they did not recall how to unselect the wrong category, the researcher asked them if they remembered how and then reminded them how to do this if not. The same researcher worked with all participants to reduce the potential for inconsistency.

Analysis

For search tasks, each child completed three type A tasks and three type B tasks with each interface. These tasks were averaged by grade (1, 3, and 5) and task type (A and B) for analysis. All time data was analyzed using repeated measures analysis of variance (ANOVA) and is graphed with 95% confidence intervals. Preliminary analysis indicated no effect by task set, so this factor was excluded. Post-hoc analyses were done with Tukey tests to explore significant results in factors with more than two levels. For subjective questions (e.g. likeability) and counts (e.g. hints) for which assumptions of normality may not hold, non-parametric statistical tests were used. All analyses were performed overall and by grade. Some non-significant results by grade may be due to lack of power with only 12 participants per grade.

Results: Simple Study

Browsing tasks

During the browsing tasks, we counted the number of Boolean queries a child created in each interface. Although the search tasks in the simple study were all one-category searches, there was nothing to prevent children from creating multi-category, Boolean queries while browsing. A Boolean query was defined as having at least two leaf-level categories selected. Overall, children in all 3 grades created 183 Boolean queries in the flat interface and 79 in the hierarchical. To analyze whether there was a difference by interface, we conducted a Wilcoxon signed rank test. The results of this test were significant ($p<0.01$), indicating that significantly more Boolean queries were created in the flat interface overall. The same was also true for each individual grade. First graders created 47 Boolean queries in the flat interface and 23 in the hierarchical. Third graders created 59 Boolean queries in the flat interface and 21 in the hierarchical. Fifth graders created 77 Boolean queries in the flat interface and 35 in the hierarchical. All of these differences were significant at $p=0.02$ or less.

Combining both interfaces and looking at the data by grade, first graders created an average of 5.8 Boolean queries, third graders an average of 6.7, and fifth graders an average of 9.3. To analyze whether there was a difference by grade, we conducted a Kruskal Wallis test. The results of this test were not significant, indicating that age did not seem to be a factor in the number of Boolean queries created.

We also counted the number of children who found and used the navigation buttons on their own in each interface while they were browsing. In the flat interface, overall 9 children used the More Choices buttons and 27 did not. To see if these values differed from the null hypothesis that half the children would use the buttons, we conducted a binomial test. The results of this test were significant ($p<0.01$), indicating that significantly more children did not find this button than did find it overall. Looking at each grade individually, only first graders demonstrated significant problems finding the More Choices button. In fact, none of the first graders used the More Choices buttons. Differences in other grades were not significant. In the hierarchical interface, 21 children from all three grades used the Up Arrow and 15 did not. The results of a binomial test were not significant. However, in fifth grade, 9 children used the Up Arrow and 3 did not, a significant difference ($p=0.04$).

Combining both interfaces and looking at the data by grade, there were 4 uses of any navigation tool in first grade, and 13 each in the third and fifth grades. To analyze this data, we conducted a Fisher exact test. This difference was significant ($p<0.01$), indicating that significantly more children in the third and fifth grades used the navigation tools on their own than first grade children.

Searching tasks

Search times for all 36 children were submitted to a 3 (grade: 1/3/5) x 2 (interface: flat/hierarchical) x 2 (task type: A/B) ANOVA. Results of this analysis indicated a significant difference by grade $F(2,33)=23.99$, $p<0.01$ (Figure 8) and a significant interaction effect between interface and task type, $F(1,33)=26.64$, $p<0.01$ (Figure 9). Tukey post-hoc tests on grade indicated that fifth graders were faster than first graders, and third graders were faster than first graders.

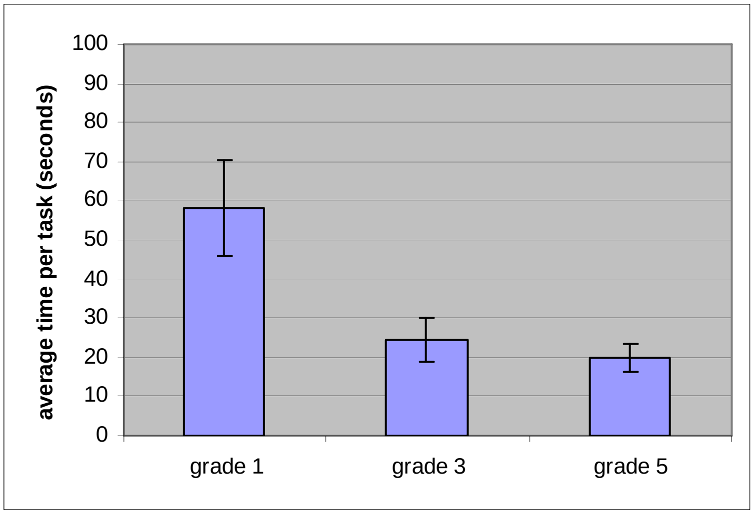

Figure 8. Mean time by grade for simple search tasks (n=36)

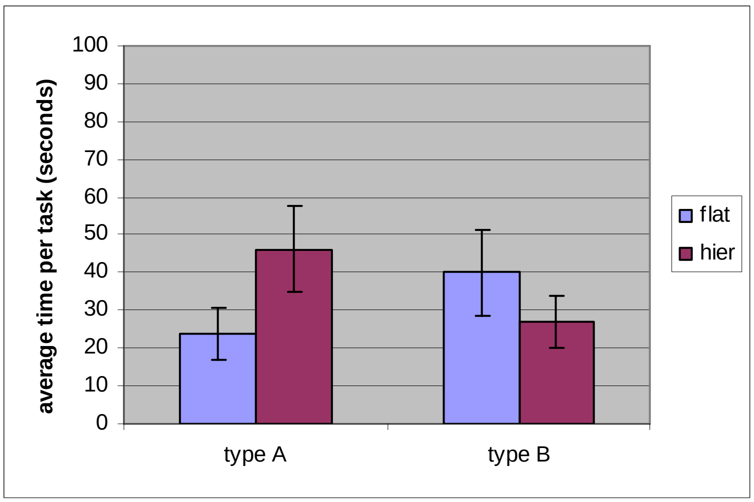

Figure 9. Mean time by interface and task type for simple search tasks (n=36)

Tukey post-hoc tests on the interaction between interface and task type indicate that type A tasks in the flat interface were faster than type A tasks in the hierarchical interface and type B tasks in the flat interface. Type B tasks in the hierarchical interface were also faster than type A tasks in the hierarchical interface. We expected to find no difference between type A and type B tasks in the hierarchical interface since they required the same number of steps to complete in that interface. However, the type A task sets contained a task that involved the Format category, a relatively abstract term. Children took more than twice as long and required more than twice as many hints to complete the task with this category in the hierarchical interface compared to the averages of the other tasks done with the hierarchical interface.

At the individual grade level, search times for all 12 children were submitted to 2 (interface) x 2 (task type) ANOVAs. In each grade, significant interaction effects were found between interface and task type with p values of 0.03 or less. In first grade, Tukey post-hoc tests indicated that type A tasks in the flat interface were faster than type A tasks in the hierarchical. In third grade, none of the post-hoc tests reached significance. In fifth grade, type A tasks in the flat interface were faster than type A tasks in the hierarchical and type B tasks in the flat.

In addition to timing children on tasks, we also counted how many hints they needed to complete the tasks. In the flat interface, children needed 79 total hints, compared to 84 in the hierarchical. To analyze whether there was a difference by interface, we conducted a Wilcoxon signed rank test. The results of this test were not significant, indicating that there was no difference between the two interfaces. The same was true for each individual grade. Combining both interfaces and looking at the data by grade, first graders needed an average of 10.1 hints, third graders an average of 2.25, and fifth graders an average of 1.25. To analyze whether there was a difference in the number of hints by grade, we conducted a Kruskal Wallis test. This test was significant ($p<0.01$), indicating that younger children needed more hints. Looking at the hints in the flat interface by task type (type A vs. type B), a Wilcoxon signed rank test was not significant, indicating that the paging required for type B tasks did not result in the need for more hints. The same was also true for hints in the flat interface for each grade.

Subjective tasks

For interface difficulty, children were asked how hard an interface was to use and given three choices: hard, medium, and easy (Table 1). For interface likeability, children were asked if they liked using an interface and were given three choices: not much, a little, and a lot (Table 2). We assigned these choices values from 1 to 3 and conducted Wilcoxon signed rank tests to compare their answers. The results of these tests were not significant overall or within any grade for either difficulty or likeability, indicating no difference between the two interfaces. After the children had used both interfaces, they were asked to choose whether they liked either interface better or both the same. Of the children, 11 preferred the simultaneous interface, 5 preferred the sequential, and 20 liked both equally (Table 3). We ordered these choices with “both” as the median value and performed a one-sample median test. The results of this test were not significant, indicating neither interface was strongly preferred.

	Hard (flat/hierarchical)	**Medium (flat/hierarchical)**	**Easy (flat/hierarchical)**
Grade 1	1/1	2/3	9/8
Grade 3	0/0	5/6	7/6
Grade 5	0/1	3/4	9/7
Overall	1/2	10/13	25/21

Table 1. Interface difficulty for simple study

	Not much (flat/hierarchical)	**A little (flat/hierarchical)**	**A lot (flat/hierarchical)**
Grade 1	0/1	1/2	11/9
Grade 3	0/0	2/2	10/10
Grade 5	0/0	2/4	10/8
Overall	0/1	5/8	31/27

Table 2. Interface likeability for simple study

	Flat	Both	Hierarchical
Grade 1	1	10	1
Grade 3	4	6	2
Grade 5	6	4	2
Overall	11	20	5

Table 3. Interface preference for simple study

Results: Boolean Study

Browsing Tasks

While browsing, children created 224 Boolean queries in the flat interface and 104 in the hierarchical interface. The results of a Wilcoxon signed rank test were significant ($p<0.01$), indicating that significantly more Boolean queries were created in the flat interface. At the individual grade level, a similar pattern emerged, though the differences were not always significant. First graders created 60 Boolean queries in the flat interface and 34 in the hierarchical. This difference was not significant. Third graders created 62 Boolean queries in the flat interface and 26 in the hierarchical. This difference was borderline significant ($p=0.06$). Fifth graders created 102 Boolean queries in the flat interface and 44 in the hierarchical. This difference was significant ($p=0.02$). Combining both interfaces and looking at the data by grade, first graders created an average of 7.9 Boolean queries, third graders an average of 7.3, and fifth graders an average of 12.2. To analyze whether there was a difference by grade, we conducted a Kruskal Wallis test. The results of this test were not significant, indicating that age did not seem to be a factor in the number of Boolean queries created.

We also counted the number of children who used navigation buttons on their own. In the flat interface, 12 children used the More Choices buttons, while 24 did not. The results of a binomial test were significant ($p<0.01$), indicating that significantly more children did not find this button than did find it. A similar pattern emerged in first and third grade, though the difference wasn't always significant. In first grade, 1 child used the More Choices buttons and 11 did not. This difference was significant ($p<0.01$). In third grade, 4 children used the More Choices buttons and 8 did not. This difference was not significant. Finally, in fifth grade, 7 children used the More Choices buttons and 5 did not. This difference was not significant.

In the hierarchical interface, 23 children used the Up Arrow and 13 did not. The results of a binomial test were significant ($p=0.05$), indicating that significantly more children found this button than did not find it. A similar pattern emerged in first and fifth grade. In first grade, 9 children used the Up Arrow and 3 did not. This difference was significant ($p=0.04$). In third grade, 4 children used the Up Arrow and 8 did not. This difference was not significant. Finally, in fifth grade, 10 children used the Up Arrow and 2 did not. This difference was significant ($p=0.01$). Combining both interfaces and looking at the data by grade, there were 10 uses of a navigation tool in first grade, 8 in third grade, and 17 in fifth grade. To analyze this data, we conducted a Fisher exact test. This test was significant ($p=0.03$), indicating that significantly more children in fifth grade used the navigation tools on their own than third and first grade children.

Searching Tasks

Search times for all 36 children were submitted to a 3 (grade) x 2 (interface) x 2 (task type) ANOVA. Results of this analysis indicated significant differences by grade $F(2,33)=19.96$, $p<0.01$ (Figure 10) and interface $F(1,33)=53.25$, $p<0.01$, and a significant interaction effect between interface and task type, $F(1,33)=18.71$, $p<0.01$ (Figure 11). For the interface effect, the flat interface was faster than the hierarchical.

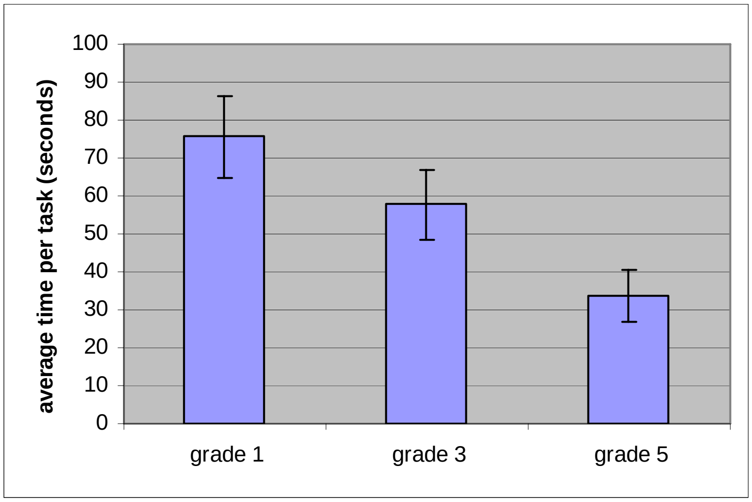

Figure 10. Mean time by grade for Boolean search tasks (n=36)

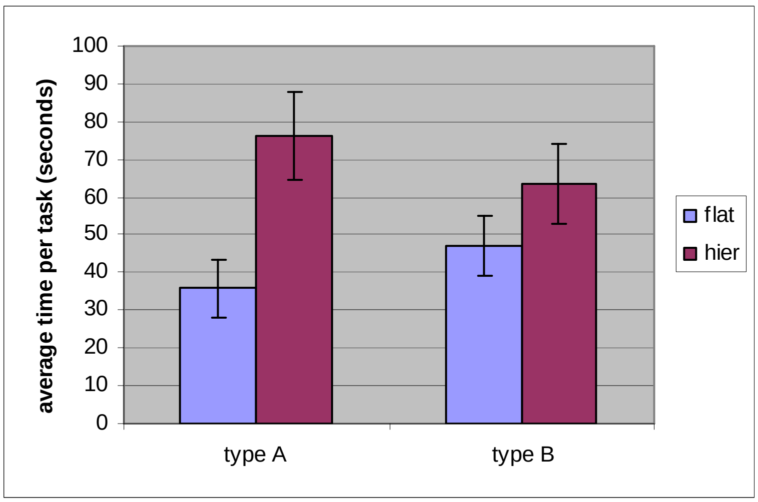

Figure 11. Mean time by interface and task type for Boolean search tasks (n=36)

Tukey post-hoc tests on grade indicated that all pairwise differences were significant: older children were always significantly faster than younger children. Tukey post-hoc tests on the interaction between interface and task type indicated that all pairwise differences were significant except for the difference between type A and type B tasks in the hierarchical interface. This result makes sense given that the flat interface was significantly faster overall, and that the type A and type B tasks in the hierarchical interface were functionally equivalent.

At the individual grade level, search times for all 12 children were submitted to 2 (interface) x 2 (task type) ANOVAs. In all three grades, there was a significant effect by interface ($p=0.02$ or less), with the flat interface faster than the hierarchical. In all three grades, there was also a significant interaction between interface and task type ($p=0.03$ or less). Tukey post-hoc tests on the interaction effect indicated a similar pattern to the overall results. Some of the pairwise differences were significant in all three grades, but the difference between type A and type B tasks in the hierarchical interface was never significant.

In the flat interface, children needed 116 total hints to complete the searching tasks, compared to 221 in the hierarchical interface. This difference was significant ($p<0.01$), indicating that significantly more hints were required in the flat interface. A similar pattern emerged at the individual grade level, but the difference was only significant in first grade. First graders needed 76 hints in the flat interface, compared with 142 in the hierarchical. This difference was significant ($p=0.01$). Third graders needed 31 hints in the flat interface and 58 in the hierarchical while fifth graders needed 9 total hints in the flat and 21 in the hierarchical. These differences were not significant. Combining both interfaces and looking at the data by grade, first graders needed an average of 18.2 hints, third graders an average of 7.4, and fifth graders an average of 2.5. A Kruskal Wallis test showed that these differences were significant ($p<0.01$), indicating that younger children needed more hints than older children. Looking at the hints in the flat interface by task type (type A vs. type B), a Wilcoxon signed rank test was not significant, indicating that the paging required for type B tasks did not result in the need for more hints. The same was also true for hints in the flat interface for each grade.

Subjective Tasks

For interface difficulty and likeability, we conducted Wilcoxon signed rank tests to compare the responses for each interface (Tables 4 and 5). The results of both of these tests were significant overall, indicating that the flat interface was considered easier and was better liked than the hierarchical interface overall. At the individual grade level, we did not find any significant differences for either difficulty or likeability. For overall interface preference, 19 children preferred the simultaneous interface, 4 preferred the sequential interface, and 13 liked both equally (Table 6). The results of a one-sample median test were significant ($p<0.01$), indicating that more children preferred the flat. At the individual grade level, only third graders expressed a significant preference ($p<0.01$), also in favor of the flat.

	Hard (flat/hierarchical)	**Medium (flat/hierarchical)**	**Easy (flat/hierarchical)**
Grade 1	1/3	2/5	9/4
Grade 3	0/1	4/9	8/2
Grade 5	0/1	1/5	11/6
Overall*	1/5	7/19	28/12

Table 4. Interface difficulty for Boolean study (* indicates significant difference)

	Not much (flat/hierarchical)	**A little (flat/hierarchical)**	**A lot (flat/hierarchical)**
Grade 1	0/1	2/3	10/8
Grade 3	0/0	1/3	11/9
Grade 5	0/1	0/4	12/7
Overall*	0/2	3/10	33/24

Table 5. Interface likeability for Boolean study (* indicates significant difference)

	Flat	Both	Hierarchical
Grade 1	3	8	1
Grade 3*	8	4	0
Grade 5	8	1	3
Overall*	19	13	4

Table 6. Interface preference for Boolean study (* indicates significant difference)

Comprehension Tasks

In the Boolean study, children completed two additional tasks where they were asked to select two category buttons and then to explain what kinds of books they had found. The goal of these tasks was to elicit whether or not the children understood what kind of search was being created when they selected multiple categories. We decided to analyze the second of these tasks, asked after the children had completed the timed searching tasks. In the flat interface, we found that overall, 22 children understood they were creating conjunctive Boolean queries in the flat interface, compared with 14 who did not. The results of a binomial test were not significant, indicating that children were as likely to understand as not. At the individual grade level, all the fifth graders understood, a significant difference. In first grade, half the children understood and half did not and in third grade, 4 children understood and 8 did not. Neither of these differences was significant.

In the hierarchical interface, we found that overall, 18 children understood what type of task they were doing and 18 did not, indicating that children were as likely to understand as not. At the individual grade level, 11 out of 12 fifth graders understood, a significant difference. In first grade, 4 children understood and 8 did not. This difference was not significant. In third grade, 3 children understood and 9 did not. This difference was significant, indicating that children in this grade were less likely to understand. To see if there was a difference in understanding by grade, we combined both interfaces together and conducted a Fisher exact test. There were 10 instances of understanding in either interface in first grade, 7 in third grade, and 23 in fifth grade. This test was significant ($p<0.01$), indicating that fifth graders understood they were creating conjunctive Boolean queries more often than first and third graders.

Discussion

Browsing Tasks

In both studies, the browsing tasks seemed to favor the flat interface over the hierarchical interface with regard to children's ability to create Boolean queries, supporting our hypothesis (H1). During browsing, children created significantly more Boolean queries using the flat interface, both overall and within some of the grades. We did not record whether or not children understood that they were creating a Boolean query at that time, but these results suggest that creating a Boolean query was easier in the flat interface.

In both studies, children were more likely than not to find the Up Arrow button in the hierarchical interface, but less likely to find the More Choices buttons in the flat interface. Our concern about paging with the More Choices button in the flat interface was thus confirmed. We observed that children did not have any problem understanding or using the More Choices buttons once the researcher pointed them out, nor did they require more hints to complete the type B searching tasks involving paging, but this was a significant usability problem for children when they did not receive prior instruction.

Searching Tasks

In both studies, the results indicated that younger children require more time and more hints than older children to complete a directed searching task, regardless of the interface presentation or the task difficulty. For simple searches, neither interface was faster or required fewer hints. However, our second hypothesis (H2) that simple, type A tasks would be faster in the flat interface than in the hierarchical interface was supported overall and within first and fifth grade. For Boolean searches, the flat interface was faster overall and within all three grades, and required fewer hints overall and within first grade, supporting our third hypothesis (H3). The timing and hint data from both studies suggest that the concerns we had about visual overload and scanning in the flat interface were thus less serious than the problems identified in the hierarchical interface.

Subjective Tasks

For simple searches, neither interface was liked more, easier to use, or preferred, either overall or within any of the three grades. However, for Boolean searches, the flat interface was liked more, easier to use, and preferred overall, supporting our fourth hypothesis (H4). Most differences between individual grades did not reach significance in the Boolean study, but third graders did significantly prefer the flat interface. Some of these differences at the individual grade level leaned toward an advantage for the flat interface, but there may not have been enough participants to find a statistical difference.

Comprehension Tasks

In the Boolean study, older children understood that they were creating a conjunctive Boolean query more often than younger children, supporting our fifth hypothesis (H5). However, neither interface was more likely to support this understanding than not. Our sixth hypothesis (H6) that the flat interface would better support this understanding was not supported.

DESIGN GUIDELINES

The results of these studies suggest several guidelines that we believe will be useful for other designers of searching and browsing interfaces for elementary-age children.

1. *Both flat, simultaneous and hierarchical, sequential designs are useful for single-item searches*. Our research confirms that children between the ages of 6 and 11 are able to use category browsers to search for information, so these tools continue to be a reasonable way to support children's searching and browsing. For interfaces that only support selecting a single category, both a two-level sequential presentation of a hierarchical structure and a two-page simultaneous presentation of a flat structure are appropriate. However, if all of the categories fit on a single page, the flat presentation is likely to be faster.

2. *Flat, simultaneous designs better support conjunctive Boolean searching and browsing*. We suggest that if Boolean queries are supported, a simultaneous presentation of a flat structure is preferable to a sequential presentation of a hierarchy, at least for conjunctive queries between facets. During free browsing tasks, children spontaneously created more conjunctive Boolean queries in the flat interface. In directed searches, children were also faster, required fewer hints, and considered creating conjunctive Boolean queries faster, easier, and preferable in the flat interface.

3. *Boolean search must be implemented with care.* Not all elementary-age children, particularly those in first and third grades, understand Boolean functionality, so the interface needs to work regardless of whether children understand the logic. Features such as automatic creation of Boolean queries, the constant presence of items to choose from in a results section on the same screen, and the prevention of no-hit searches by graying out useless categories, can all make the interaction easy even if a child does not understand the underlying Boolean logic. Despite the fact that many of the younger children did not understand that they were creating Boolean queries, they created about as many such queries while browsing spontaneously as the older children.

4. *Navigation tools must be large and prominent.* Finally, our research suggests that young children are comfortable navigating a two-level hierarchy with a large, conspicuously placed navigation arrow. On the other hand, smaller, less obviously placed buttons for paging in a flat presentation may be too small and/or get lost among other buttons. As a result, we recommend larger, more prominently placed paging buttons, or avoiding paging altogether by decreasing the number of categories.

CONCLUSIONS AND FUTURE WORK

We presented the results of two studies designed to improve the understanding of elementary-age children's searching and browsing skills using two common category browser interfaces in a children's digital library. Our results indicate that a flat category structure presented simultaneously has a number of advantages over a hierarchical category structure presented sequentially for searching and browsing activities that involve conjunctive Boolean queries. For simple, one-category queries, the two designs appear to be equally suitable. We believe these results will be applicable to other searching and browsing interfaces for children, such as digital libraries and search engines. Any large collection of data will require conjunctive queries to narrow down the result set to a reasonable size, and we presented design guidelines to help other designers support this functionality.

In the future, we would like to see research undertaken to explore five other areas. First, our research compared only two of the three possible structure/presentation combinations in a category browser. We compared a flat, simultaneous interface to a hierarchical, sequential interface. We did not look at a hierarchical, simultaneous design. A flat, sequential design is not possible. As a result, we were not able to determine if either structure or presentation had more of an impact on performance and preference. Our hypothesis is that a hierarchical, simultaneous design, such as is implemented in Microsoft Windows Explorer, would be more difficult for children to use than either of the other two designs because it would have many items on screen and require a lot of navigation. Future studies might explore this question, and whether changes to structure or presentation have a greater effect.

Second, our research evaluated only conjunctive Boolean search between facets. We chose to support conjunction because a number of studies indicate that children have an easier time with conjunction than disjunction (Neimark & Slotnick, 1970), and conjunction narrows down the search results to a more manageable size. We chose to evaluate searching only between facets because certain within-facet searches are not possible in our browser. For example, within the Length category, a book cannot be both Long and Short, but within the Color category, it is possible for a book to be both Red and Blue. Within-facet searches are likely to be easier than between facet searches in a hierarchical design since no backtracking is necessary. It may be useful to evaluate this and other logical combinations in the future.

Third, we would like to see more study on children's browsing skills. In our studies, strict time limits were not enforced during browsing tasks and children were not asked about comprehension while browsing. Further studies might reveal more information about children's comprehension of Boolean queries while browsing. Fourth, we would like to see more work done on evaluating different ways of supporting both paging and Boolean comprehension, the two areas where our new design presented challenges. Finally, we would like to see studies done with more children at multiple grade levels to ensure sufficient power to detect differences within each grade. A number of our within-grade measurements suggested more advantages for the flat interface for Boolean queries, but did not reach significance.

ACKNOWLEDGMENTS

We would like to thank the entire ICDL research team, the Columbia Association, and the Hillcrest, Northfield, Clemens Crossing, and Holy Trinity elementary schools for making this research possible. This research was partially funded by a National Science Foundation ITR grant, #0205082, and an Institute of Museum and Library Services National Leadership grant, #LG02020026-02. Any opinions, findings, conclusions, or recommendations expressed in this material do not necessarily reflect the views of these funding organizations.

REFERENCES

Ahlberg, C. & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. *Proceedings of Human Factors in Computing*, 313-317.

Borgman, C., Hirsh, S., Walter, A., & Gallagher, A. (1995). Children's searching behavior on browsing and keyword online catalogs: The Science Library Catalog project. *Journal of the American Society for Information Science and Technology, 46 (9)*, 663-684.

Busey, P. & Doerr, T. (1993). Kid's Catalog: An information retrieval system for children. *Youth Services in Libraries, 7 (1),* 77-84.

Corporation for Public Broadcasting (CPB) (2002). Connected to the future: A report on children's internet use. Retrieved June 28, 2005 from http://www.cpb.org/stations/reports/connected.

Day, J., Janus, A., & Davis, J. (2005). Computer and internet use in the United States: 2003. *U.S. Census Bureau*, October 2005.

Demunter, C. (2005). The digital divide in Europe. *Eurostat 38/2005*, ISSN 1561-4840.

Druin, A. (2005). What children can teach us: Developing digital libraries for children. *Library Quarterly, 75 (1)*, 20-41.

Druin, A., Bederson, B., Hourcade, J., et al. (2001). Designing a digital library for young children: An intergenerational partnership. *Proceedings of the Joint Conference on Digital Libraries,* 398-405.

Druin, A., Bederson, B., Weeks, A., et al. (2003). The International Children's Digital Library: Description and analysis of first use. *First Monday, 8 (5).*

Edmonds, L., Moore, P, & Balcom, K. (1990). The effectiveness of an online catalog. *School Library Journal, 36 (10),* 28-33.

Hochheiser, H. & Shneiderman, B. (2000). Performance benefits of simultaneous over sequential menus as task complexity increases. *International Journal of Human-Computer Interaction*, 12(2), 173-192.

Hourcade, J., Bederson, B., Druin, A., & Guimbretiere, F. (2004). Differences in pointing task performance between preschool children and adults using mice. *ACM Transactions on Computer-Human Interaction, 11 (4)*, 357-386.

Hutchinson, H., Bederson, B., & Druin, A. (2006). The evolution of the International Children's Digital Library searching and browsing interface. *Proceedings of Interaction Design and Children*, 105-112.

Inkpen, K. (2001). Drag-and drop versus point-and-click mouse interaction styles for children. *ACM Transactions on Computer-Human Interactions, 8 (1),* 1-33.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109 (3),* 490-501.

Kids Catalog Web (kcweb.tlcdelivers.com).

Kragler, S. & Nolley, C. (1996). Student choices: Book selection strategies of fourth graders. *Reading Horizons, 36 (4),* 354-365.

Kuhlthau, C. (1988). Meeting the information needs of children and young adults: Basing library media programs on developmental states. *Journal of Youth Services in Libraries,* Fall 1988, 51-57.

Livingstone, S. (2003). Children's use of the internet: Reflections on the emerging research agenda. *New Media & Society, 5 (2)*, 147-166.

Marchionini, G. & Teague, J. (1987). Elementary students' use of electronic information services: An exploratory study. *Journal of Research on Computing in Education 20(2),* 139-155.

Miller, D. (1981). The depth/breadth tradeoff in hierarchical computer menus. *Proceedings of the Human Factors Society*, 296-300.

Mills, C. & Weldon, L. (1987). Reading text from computer screens. *ACM Computing Surveys, 19 (4),* 329-358.

Moore, P. & St. George, A. (1991). Children as information seekers: The cognitive demands of books and library systems. *School Library Media Quarterly, 19*, 161-168.

National Telecommunications and Information Administration (NTIA) (2004). A nation online: Entering the broadband age. Retrieved June 28, 2005 from http://www.ntia.doc.gov/reports/anol/index.html

Neimark, E. & Slotnick, N. (1970). Development of the understanding of logical connectives. *Journal of Educational Psychology, 61 (6*), 451-460.

Pejtersen, A. (1989). A library system for information retrieval based on a cognitive task analysis and supported by an icon-based interface. *ACM Conference on Information Retrieval*, 40-47.

Reuter, K. & Druin, A. (2004). Bringing together children and books: An initial descriptive study of children's book searching and selection behavior in a digital library. *Proceedings of the 67th Annual Meeting of the American Society of Information Science and Technology,* 339-348.

Revelle, G., Druin, A., Platner, M., et al. (2002). A visual search tool for early elementary science students. *Journal of Science Education and Technology, 11(1)*, 49-57.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382-439.

Schneider, K. (1996). Children and information visualization technologies. *Interactions, 3 (5),* 68-74.

Soloway, E., Guzdial, M., & Hay, K. (1994). Learner-centered design: The challenge for HCI in the 21st century. *Interactions, April 1994*, 36-48.

StoryPlace (www.storyplace.org).

Strommen, E. (1994). Children's use of mouse-based interfaces to control virtual travel. *Proceedings of Human Factors in Computing Systems*, 405-410.

Thomas, J. (1980). Acquisition of motor skills: Information processing differences between children and adults. *Research Quarterly for Exercise and Sport, 51 (1)*, 158-173.

Tversky, B. (1985). Development of taxonomic organization of named and pictured categories. *Developmental Psychology, 21*(6), 1111-1119.

Tversky, A. & Kahneman, D. (1974). Judgments under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Wendelin, K. & Zinck, R. (1983). How students make book choices. *Reading Horizons, 23,* 84-88.

Yee, K., Swearingen, K., Li, K. & Hearst, M. (2003). Faceted metadata for image searching and browsing. *Proceedings of Human Factors in Computing Systems*, 401-408.