# Relationship Learning Software: Design and Assessment

Kyla A. McMullen and Gregory H. Wakefield

The University of Michigan, Ann Arbor MI 48109-2121, USA
kyla@umich.edu, ghw@umich.edu

**Abstract.** Interface designers have been studying how to construct graphical user interfaces (GUIs) for a number of years, however adults are often the main focus of these studies. Children constitute a unique user group, making it necessary to design software specifically for them. For this study, several interface design frameworks were combined to synthesize a framework for designing educational software for children. Two types of learning, relationships and categories, are the focus of the present study because of their importance in early-child learning as well as standardized testing. For this study the educational game Melo's World was created as an experimental platform. The experiments assessed the performance differences found when including or excluding subsets of interface design features, specifically aesthetic and behavioral features. Software that contains aesthetic, but lack behavioral features, was found to have the greatest positive impact on a child's learning of thematic relationships.

**Keywords:** human computer interaction, educational technology, interactive systems design, user interface design.

## 1 Introduction

Interface designers have been studying how to construct graphical user interfaces (or GUIs) more suitable for humans for a number of years. However, these designers tend to focus solely on issues concerning adults and tend to neglect those that must be considered when designing GUIs for children. The capabilities of children differ from adults in several areas including, motor skills, literacy levels, and attention spans. Because of these differences, the study of interface design for children deserves as much attention as the study of interface design for adults.

In particular, we would like to study the design of user interfaces used in educational software for children four to six years of age. The ages of four years old to six years old are the most important years of learning for children [13]. We would like to study how to most effectively design software to aid learning during these critical years. Typically the designers of educational software for children focus on aesthetics, such as flashy graphics or intriguing sounds, in order to hold the attention of the child playing the game. Although these aesthetics

J.A. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2009, LNCS 5612, pp. 631–640, 2009.

and graphics hold the child's attention, there is no empirical evidence that learning is actually taking place. Educational software designers should perform more empirical analysis to ensure that the game interfaces that are created for children actually facilitate learning. Though the interest in interactive multimedia continues to grow, so far its design has been primarily driven by technological advances, rather than by theoretical principles [8], [14].

The goal of this study is to evaluate the effects of user interfaces for children's learning technology, in order determine how to best facilitate learning, specifically, relationship learning. Simple Adobe Flash based games were used to develop the educational software necessary to conduct the study. They can be run on any computer with a free Flash Player installed. We hypothesize that an educational game interface that has behavioral components (has an easy to understand interface, gives hints and clues, responds to every action performed by the child, etc) and that has aesthetic components (bright colors, soothing sounds, etc) will facilitate learning in an educational software environment better than an educational game interface lacking one of these components.

## 2 Background and Significance

### 2.1 Child Centered Interface Design

This study seeks to improve the quality of educational technology design for children. Interface designers sometimes forget that children are young people that constitute an entirely different computer user population with their own culture, norms, and complexities [2]. Most research on interface design focuses on adults as the primary users, neglecting the differences that children may have when interacting with educational technology.

When designing software for young children, designers should focus on a particular age range, because children of different ages have vastly different preferences and levels of skills [7]. The present study focuses on the four to six year old age group. As stated earlier, these are the most formative years for a child's learning. The focus lies here because the research on learning technologies for children has been conducted primarily using software for older children. Furthermore, the existing guidelines for designing learning technology for children do not distinguish between children of different age groups.

General guidelines for designing materials for learning technology for children have been postulated (e.g. Jones [9], Clanton [5], Neilsen and Molich [11], Norman [12], and Buckleitner [3]). These necessary design features are also pictured in Figure 1. All of these interface design features can be grouped into 4 following categories: Aesthetic, Behavioral, Interaction, and Uncertainty. The choices made in designing the interface for Melo's World was derived directly from these guidelines.

### 2.2 Relationship Learning

The game that was created for this study focuses on enriching relationship learning skills in children. As we know, objects in the world can be related in a

| Interface Design Features | | | |
|---|---|---|---|
| Aesthetic features | bright and highly saturated hues | Behavioral features | clear goals and an established quest |
| | smiling faces | | continual and immediate feedback |
| | soothing and harmonious sounds | | feedback that uses sound and graphics |
| | simple melodies | | messages personally addressing the user |
| | rounded and/or symmetrical objects | | messages that use everyday English |
| | rhythmic percussive beats | | goals for each level of game play |
| | humor | | varied feedback responses |
| Interaction | obvious interface | | hints |
| | well organized concepts | | provides shortcuts |
| | independent exploration | | makes failure fun |
| | familiar objects | | various interaction options |
| | dynamic features | | re-do and undo options |
| | appropriate feedback | Uncertainty | different types of problems |
| | simple design | | challenging yet completable tasks |
| | opportunities for trial and error | | spread out clues tools and obstacles |
| | media rewards | | uncertain outcome |
| | simple help and documentation | | hidden information |
| | lets user exit at any time | | |
| | error prevention and recovery | | |

**Fig. 1.** Interface Design Features

multitude of ways. For example, objects can be related by the ways in which the objects participate in the same event or theme (cats eat mice, people read books, birds build nests). These assorted external relations between objects are referred to as thematic relationships. Research has shown that thematic groupings form the basis of children's categorization and learning [10]. Wyche has shown that children in low socioeconomic environments tend to struggle with relationship learning and categorization stills [16]. For those interested in lowering the socio-economic status barriers often found in educational software, this study is of particular interest. These skills are critical in order for children to perform well in school. Classification and categorization of objects are also a major skill set that is necessary to perform well on standardized testing. Often a school's standardized testing performance is used to determine the amount of funding that is received; with the lower performing school systems receiving less money than the higher performing school systems. Relationship learning helps to form the foundation upon which children learn to categorize.

## 2.3 Categorization and Classification

There is evidence to suggest that the way in which participants categorize entities externally reflects their internal, mental representation of these concepts. Card sorting is a common empirical technique to externally elicit a representation of a person's internal categorical structure. Card sorting originated in George Kelly's Personal Construct Theory. Card sorts have been used historically in order to elicit each individual's individual and often semi-tacit, understanding about objects world and their relationships to one another. Eliciting the structures of knowledge via card sorts is a more reliable indicator of a user's expertise and learning than quantities of facts, as demonstrated by a series of investigations by Chi and Koeske [4].

This study used the free card sorting (or open card sorting) method in order to assess the participants' initial categorical structure. In free card sorts, neither categories nor criteria are specified in advance and the user may arrange the cards in as many groups as they wish. In order to compare the user's pre-existing categorical structure, as evidenced by the groups formed during the free-sort, to the correct categorical structure, a comparison metric had to be used. The metric of choice is called the Card Sort Edit Distance (CSED) metric which will later be described in more detail.

## 3 Melo's World

For this study, we created an educational software game in Adobe Flash called Melo's World. By creating and using Melo's World, we were able to easily manipulate the game's design features. There were three versions of Melo's World created. All three versions contain the base features. The Aesthetic-only version contains all of the base features along with the aesthetic features. The Behavioral-only version contains all of the base features along with the behavioral features. Finally, the Aesthetic+Behavioral version contains all of the base features along with the aesthetic and behavioral features. The divisions of these features can be seen in Figure 2.

At the beginning of game play, the child is able to choose the specific portion of the game that they would like to explore as seen in Figure 3. In the first scene, the child is shown a map of Melo's World where they are allowed to pick the level they would like to play. Each level of the game focuses on enriching the child's relationship learning skills in some familiar context (a messy room, school, the playground, grandma's house, and the grocery store).

| **Interface Design Features** | |
|---|---|
| **Base features** | obvious interface |
| | well organized concepts |
| | independent exploration |
| | familiar objects |
| | dynamic features |
| | appropriate feedback |
| | simple design |
| | opportunities for trial and error |
| | media rewards |
| **Aesthetic features** | Bright and highly saturated hues |
| | smiling faces |
| | soothing and harmonious sounds |
| | simple melodies |
| | rounded and/or symmetrical objects |
| | rhythmic percussive beats |
| **Behavioral features** | clear goals and an established quest |
| | continual and immediate feedback |
| | feedback that uses sound and graphics |
| | messages personally addressing the user |
| | messages that use everyday English |
| | goals for each level of game play |
| | varied feedback responses |
| | hints |

**Fig. 2.** Interface Design Features used to Develop Melo's World

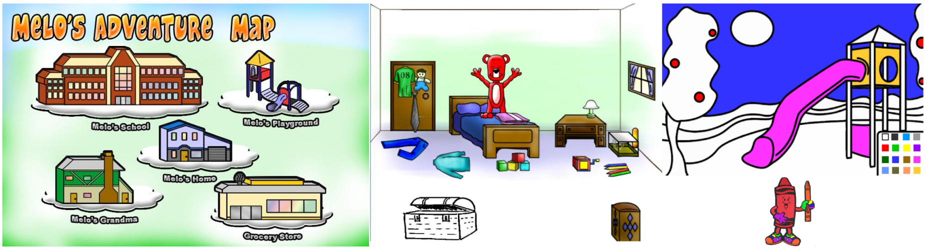


**Fig. 3.** Scene 1: Adventure Map - In this scene, the user is able to select the specific game module that they would like to play. Scene 2: Melo's Messy Room - In this scene, the child is presented with Melo's Messy Room. It is the child's job to help pick up all of the objects in the room. Scene 3: Coloring Book - This is the media reward that the child receives for correctly classifying all of the items in the room.

In order to simplify the experiment, the child was only allowed to choose the module called Melo's Home. This module consists of 3 scenes as shown in Figures 3. After choosing to explore Melo's home, the child is brought to the next scene which is called Melo's Messy Room. In this room, there are a number of items messily scattered about the floor and door. All of these items fall into the clothing or toys category. Along with the clothing and toys are a clothes hamper and a toy-box. In this scene, the child must clean up the room by correctly putting away all of the items strewn about the room by dragging each item from the floor and into its proper place. All of the toys belong in the toy-box and all of the clothes belong in the clothes hamper.

Following the game's successful completion, the child is given a media reward for playing the game. This is delivered in the form of a coloring book (as seen in Figure 3) activity. In the coloring book activity, the child is initially presented with a black and white picture, and is allowed to color the scene in any manner that they would like. After they have indicated that they are finished the coloring portion, the child is brought back to the game's initial screen.

## 4 Experiment 1

### 4.1 Procedure

After first obtaining written parental and verbal participant consent, the participants were tested individually in a small room containing a table, a laptop with an external mouse, 2 chairs, and a video camera placed on a tripod behind the child. First, the Flashcard Pretest Task was performed. In this task, the stimuli presented were 9 black and white images on 8.5 x 11 cards. Each flashcard contained one image of an object that the child would encounter while playing the software. The images on the flashcards fell into one of two categories: toys (blocks, crayons, doll, jack-in-the-box) and clothing (jeans, jersey, pants, sweater, and tie). At the start of the task, the participants were given a stack of flashcards that were randomly arranged. Each child was given the instructions:

*Put these flashcards into groups of things that go together.* After the child had grouped the flashcards, the researcher noted the groupings that were made.

Next, each child played the educational computer game, corresponding to their randomly assigned condition (if the child was assigned the control condition, the game was not played). Each child played the game for two iterations. Finally, each participant performed the Flashcard Task posttest, using the same method mentioned described above in the Flashcard pretest task.

### 4.2 Participants

The participants were fifty-five four to six year old children (29 female and 26 male)from various after school programs, day care centers, kindergarten classes, and Head Start centers in the Ypsilanti and Ann Arbor Michigan areas. All of these children had received between one and two years of formal schooling. The mean age was 5.01 years.

### 4.3 Evaluation Measures

First, each child's performance was measured by evaluating the differences between their pretest and posttest scores in a free sorting Flashcard task. We also assess the child's performance as it changes while playing the game.

The Flashcard Tasks (the pretest and the posttest) were evaluated, using the CSED metric [6] mentioned earlier. The CSED metric is a measure of how many cards need to be moved in order to transform one card sort into another. The CSED metric was used to measure how many cards are necessary to move in order to convert the participant's sorted groups into the correctly sorted groups. An integer score between 0 and 7 can be obtained using the CSED metric. A score of 0 indicates there were no cards that needed to be moved, and the correct sorting had been produced. A score of 7 indicates every card needed to be moved in order to transform the card sort create by the participant into the correct card sort. Once this score had been generated for the pretest and the posttest, the scores were compared to each other. Using the CSED metric helped to determine how well the participant was able to sort the cards into thematic categories. Furthermore, it quantifies the effect that the software had on the child's learning of categories, from playing the game.

While each participant played the software, video data was collected(from those whose parents had consented video recording). From the video, the number of trials that each participant needed to make in order to correctly classify each item in the room (as either a toy or a piece of clothing) was recorded for each round of game play. This metric produces an integer score between 9 and 18 (assuming that an item could be incorrectly classified only once). A total number of trials closer to 9 would indicate that the child possessed the skill necessary to correctly classify each item. A number of necessary trials that was closer to 18 would indicate that the child had incorrectly classified each item before selecting the correct classification.

**Multiple Comparisons**

Dependent Variable: difference
LSD

| (I) version | (J) version | Mean Difference (I-J) | Sig. |
|---|---|---|---|
| B+A | A | -1.62698* | .043 |
| | B | -.60476 | .375 |
| A | B+A | 1.62698* | .043 |
| | B | 1.02222 | .189 |
| B | B+A | .60476 | .375 |
| | A | -1.02222 | .189 |

*. The mean difference is significant at the .05 level.

**Paired Samples Test**

| | | Paired Differences | | | |
|---|---|---|---|---|---|
| | | t | Sig. (2-tailed) | Mean | Std. Error Mean |
| Pair 1 | PretestBA - PostestBA | -.159 | .876 | -.07143 | .45045 |
| Pair 2 | PretestA - PosttestA | 2.800 | .023 | 1.55556 | .55556 |
| Pair 3 | PretestB - PosttestB | 1.035 | .318 | .53333 | .51517 |

**Fig. 4.** Paired Samples t-test and ANOVA - Here a paired sample t-test is performed to determine whether there is a significant difference between the pretest scores and the post test scores for each condition. The conditions are abbreviated to the first letter of the condition name.

### 4.4 Results

Using the CSED metric to measure the performance during the flashcard task free sorting, the mean pretest score for the Behavioral+Aesthetic condition was 4.29 and the average posttest score was 4.36. In the Aesthetic-only condition, the mean pretest score was 4.56 and the mean posttest score was 3.00. In the Behavioral-only condition, the mean pretest score was 4.20 and the mean posttest score was 3.67. In the control condition the mean pretest score was 4.47. The children in the control condition only participated in the flashcard task pretest, and did not play the educational software game, nor perform the posttest. In order to determine whether there was a significant difference between the pretest scores and the post test scores for each condition (excluding the control condition), a paired samples t-test was performed. The results can be seen in Figure 4.

In comparing the Behavioral+Aesthetic condition's pretest and postetst scores, we obtain a paired difference significance value of 0.876. When the Aesthetic-only condition's pretest and postetst scores are compared, we obtain a paired difference significance value of 0.023. When the Behavioral-only condition's pretest and postetst scores are compared, we obtain a paired difference significance value of .318. From this, we can see that there is only a significant difference pretest and posttest scores in the Aesthetic-only condition.

In order to determine if there was a difference between the three versions of the game, and where those differences may lie, a one way ANOVA was performed. The results are displayed in Figure 4. In this figure, one can see that the greatest difference lies between the Aesthetic-only condition and the Behavioral+Aesthetic condition with a significance of 0.043. The differences between the other pairs of conditions(Behavioral+Aesthetic and Behavioral-only; Behavioral-only and Aesthetic-only) were not found to be significant. Each child's performance while playing the game was also recorded. The gameplay data can be seen in Figure 5. It was observed that during the first round of game play, the mean number of attempts needed in order to correctly classify all

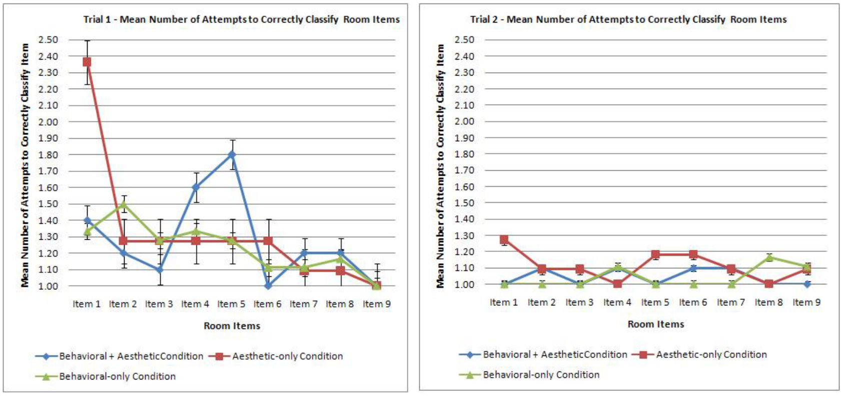


**Fig. 5.** Trial 1 and 2: Distribution of Correct Classification attempts - Along the abscissa is each room item. Item 1 is the first item correctly classified, Item 2 is the second item to be correctly classified in the room and so on. Along the ordinate is the number of incorrect categorizations observed before the item was correctly classified. Here, one can observe that although there is a steep learning curve in the Aesthetic-only condition, the performance monotonically increases, leading to less incorrect categorizations as the room items are encountered. There is no such trend for the other conditions.

of the items in the room was 11.50 for the Behavioral+Aesthetic condition, 11.91 for the Aesthetic-only condition and 11.12 for the Behavioral-only condition.

During the second round of game play, the mean number of attempts needed in order to correctly classify all of the items in the room was 9.40 for the Behavioral+Aesthetic condition, 10.00 for the Aesthetic-only condition and 9.39 for the Behavioral-only condition.

The distribution of classification attempts can be seen in Figure 5. From these figures one can see that in Trial 1, the participants in the Aesthetic-only condition required more attempts at the beginning of the game before they began to correctly classify the objects in the room. The number of attempts sharply decreases as game play continued. Although the Aesthetic-only condition had a higher initial learning curve, the number of attempts needed by each child to correctly classify the objects in the room monotonically decreased. In the Behavioral+Aesthetic condition, the average number of attempts needed in order to correctly classify the items in the room did not adhere to any specific patterns. The same can be said for the Behavioral-only condition. In the second round, the participants in all 3 game conditions behaved similarly, with the number of necessary attempts remaining close to 1.

## 5 Discussion

From the data presented in the previous section, we conclude that the Aesthetic-only version of the software produced the highest positive difference in pretest and posttest scores. The Aesthetic-only version of the software also indicates a monotonically increasing performance measure in a child's initial game play. This is very different from the expected result, that the Behavioral+Aesthetic version of the software should produce the highest positive difference in pretest and posttest scores, and that the Behavioral+Aesthetic version of the software should show the greatest improvement in performance while the child is playing the game. This finding may indicate that the addition of the behavioral components to the aesthetic components was a distraction to the child. This finding may also be due to the fact that in the aesthetic condition, the auditory instructions were excluded. As a result, the participants had to actively deduce the instructions and the goal of the game by themselves. This is a form of active learning. Active learning tasks have been found to results in greater learning than passive learning tasks [1]. The Behavioral-only condition and Behavioral+Aesthetic condition are passive learning tasks because the child is being told the necessary information, rather than having to deduce it for themselves.

## 6 Conclusion

This study sought to assess the consequences of incorporating and excluding subsets of interface design features into learning technology for children. A piece of educational software called Melo's World was developed to serve as a test facility in which we could augment the interface features and study the effects. It was found that software containing only aesthetic interface elements, thus possibly promoting active learning, produces the greatest positive differences in children learning thematic relationships. Most multimedia programs today fail because they merely add video and graphics to passive learning techniques. It does not matter whether that next page is text, graphics, or video, because the student is not doing anything. [15] These ineffective programs are using outdated, passive manners of instructing the child. In passive learning, people may absorb the facts, but they will be less active in interpreting and integrating them. Active learning results in greater learning and in more positive self-related affects and cognitions. [1] Creating educationally effective multimedia programs means taking seriously the idea of active learning. Good educational software promotes active learning, not passive learning. Furthermore, it ensures that the students are actively learning through the user interface by doing, and not simply watching.

**Acknowledgments.** I would like to thank all of the families, staff, and students at all of the kindergarten classrooms, after school programs, and head-start programs that supported this study. I would also like the Office of Naval Research - HBCU Future Engineering Faculty Fellowship Program for funding this project.

## References

1. Benware, C.A., Deci, E.L.: Quality of Learning with an Active versus Passive Motivational Set. American Educational Research Journal 21, 755–765 (1984)
2. Berman, R.A.: Preschool knowledge of language: What five-year olds know about language structure and language use. In: Pontecorvo, C. (ed.) Writing development: An interdisciplinary view, pp. 61–76. John Benjamin's Publishing, Amsteredam (1977)

3. Buckleitner, W.: The relationship between software interface instructional style and the engagement of young children. Phd Dissertation, Michigan State University (2004)
4. Chi, M.T.H., Koeske, R.D.: Network representation of a child's dinosaur knowledge. Developmental Psychology 19, 29–39 (1983)
5. Clanton, C.: An Interpreted Demonstration of Computer Game Design. In: CHI 1998 Conference Summary on Human Factors in Computing Systems, pp. 1–2 (1998)
6. Deibel, K., Anderson, R., Anderson, R.E.: Using edit distance to analyze card sorts. Expert Systems 22(3), 129–138 (2005)
7. Gelderblom, H.: Designing software for young children: theoretically grounded guidelines. In: Proceedings of the 2004 Conference on interaction Design and Children: Building A Community, pp. 121–122 (2004)
8. Hegarty, M., Quilici, J., Narayanan, N.H., Holmquist, S., Moreno, R.: Designing multimedia manuals that explain how machines work: Lessons from evaluation of a theory-based design. Journal of Educational Multimedia and Hypermedia 8, 119–150 (1999)
9. Jones, M.K.: Human-computer interaction: A design guide. Educational Technology Publications, Englewood Cliffs (1989)
10. Markman, E.: Categorization and Naming in Children. MIT Press, Cambridge (1989)
11. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people, pp. 249–256 (1990)
12. Norman, D.A.: Emotion and attractive Interactions 9(4), 36–42 (2002)
13. Palmer, J.: Pre-School Education, Pros. and Cons. A Survey of Pre-School education with Emphasis on Research Past, Present, and Future. Toronto Board of Education, Ontario (1996)
14. Park, I., Hannafin, M.J.: Empirically based guidelines for the design of interactive multimedia. Educational Technology Research and Development 41(3), 63–85 (1993)
15. Schank, R.C.: Active learning through multimedia. IEEE Multimedia 1(1), 69–78 (1994)
16. Wyche, L.G.: Conceptualization Processes in Third Grade Black Children. The Journal of Negro Education 49(4), 373–384 (1980)