# Designing Exhibits For Gender Equity

by

Toni Nicole Dancu

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Systems Science: Psychology

Dissertation Committee:
Dalton Miller-Jones, Chair
Todd Bodner
Katherine E. McDonald
Wayne Wakeland
Christine Chaillé

Portland State University

## Abstract

Gender equity has been a national and global aim for over half a century (Ceci & Williams, 2007; National Center for Education Statistics, 2003; National Science Board, 2008). While gains have been made, one area where inequity remains is spatial reasoning ability, where a large gender gap in favor of males has persisted over the years (Else-Quest, Linn, & Shibley Hyde, 2010; National Science Board, 2008; Ruble, Martin, & Berenbaum, 2006). This gender gap in spatial reasoning has had substantial societal impact on the career interests of females in areas of Science, Technology, Engineering, and Math (STEM), contributing to the larger societal need to engage non-dominant groups in these fields to reduce outsourcing (Ceci & Williams, 2007; Jaschik, 2007; Wai, Lubinski, & Benbow, 2009; White, 1992). Both spatial reasoning ability and STEM career interest have been related to science museum visits (Hamilton, Nussbaum, Kupermintz, Kerkhoven, & Snow, 1995; Salmi, 2001, 2002). However, researchers have also found a gender gap in favor of males in regard to science museum attendance and experiences once at the museum (Borun, 1999; Crowley, 2000). There are many suggestions for increasing female engagement at science museums and creating equitable experiences, but few have been systematically studied (Kekelis, Heber, & Countryman, 2005; Koke, 2005; Maher, 2005; Taylor, 2005).

This research investigated gender equitable exhibit development by enhancing a geometry exhibit with several female-friendly design features and analyzing video data to determine the effects on girls' engagement and social interactions with their caregivers. The findings suggest that incorporating several female-friendly design features leads to significantly higher engagement for girls (evidenced by greater attraction and time spent). This study also looked for any unanticipated negative effects for boys after incorporating the female-friendly design features. It is encouraging that this study was unable to detect any unintended negative effects for boys; however, such non-significant results are inconclusive and should not dissuade future research and design teams from continuing to check for unanticipated ill effects of female-friendly design features for boys. While the positive effects for girls were significant, it is important to note that they were not significantly more positive for girls than for boys; further research is needed to determine whether the female-friendly design features create a more equitable experience for girls, or a more positive experience for everyone. This study did not identify any significant differences in parent-child verbal social interactions between the two versions of the exhibit; however, the pattern of results suggests that gender discrepant parent explanations, as found by Crowley, 2001 in a children's museum, may be less of a concern for girls in science centers, providing an interesting area for future study. This research presents evidence to support incorporating female-friendly design features in future science exhibit development projects, and indicates areas where future studies are still needed to gain a deeper understanding of their effects.

**Acknowledgements**

I would like to extend my gratitude to the many people who helped make this dissertation possible and supported me throughout the process. Foremost, I would like to thank my committee for their thoughtful contributions to both my graduate education and this study. Dr. Miller-Jones, thank you for supporting me throughout this and my thesis project; thanks for always, always having my back. Dr. Bodner, thank you for providing me with such an excellent background in statistics that has already considerably impacted my career trajectory, and for reading and responding to the many rapid-fire emails I've been sending your way. Dr. McDonald, thank you for continually reviewing sections of my dissertation, for providing me with excellent dissertation examples, and for guiding me toward an understanding of coding scheme development that I will draw upon throughout my career (and knowledge of evaluation goes without saying, right?). Dr. Chaille, thank you for my most memorable course on theory, and for keeping it fresh in my mind; I look forward to working together in the museum field sometime soon. Dr. Wakeland, thank you for your patience while I worked through the application of systems theories and models to my research, and for being such a gracious host every time I visit the Harder House. I'd like to extend my gratitude to the four women in the informal learning and museum education fields who shared their expertise, lessons learned, and reading lists regarding gender: Dr. Bennie (who has been an incredible mentor throughout my graduate education), Dr. McCreedy, Ms. Koke, and Ms. Borun.

As this was an exhibit development and research project, there are a great many people at the Exploratorium who supported and contributed to this work. Eric Dimond, I could not have done this research without your collaborative efforts. I am awed by your willingness to contribute your exhibit to this study and spend months building and iterating an entirely new version for the purposes of this research; you have my deepest gratitude. Dr. Gutwill, I appreciate all of your help with this study, from conceptualization and interpretation, to your inclination to not only listen to my frustrations but also take action by troubleshooting and making it easier in any way you could. Nina, thank you for holding my hand; your emotional and professional support throughout this process has been unwavering. Kristina, I really appreciate your interest in this project, your thoughtful feedback on the exhibit, and your secret-agent label operations. Peggy, thank you for contributing your development knowledge to and securing the funds for the second version of this exhibit, and for going to bat for the final version. Mary Elizabeth, thank you for being on the development team even though you have three radical jobs, and for answering random text messages about exhibit paint color and accordions. Paul, your way with words has put me at ease about this project many more times than you know; thank you. I am also very grateful for the Illustrator crash course that made the impossible label iteration possible (and I can't wait for the follow-up CNC router course). Olympia, I cannot imagine how we would have created all of the lovely exhibit components without your keen design sense and your wiliness to work to the bone days before moving to Spain. Sebastian, thank you for noticing when I'd hit my limit of understanding with exhibit design, and taking the time to make expert observations and recommendations that truly improved the usability of the second version of this exhibit. Anne, thank you for participating on the development team, and for commiseration and celebrating in the hallways (you're nearly finished, too!). Adam, thank you for repeatedly providing your AV and computer systems expertise which made data collection possible and secure. Lisa, thank you for helping collect data from the very beginning, for filling in wherever it was needed on such short notice, and for listening to me whine in our office on a daily basis. Sarah and Lisa, thanks for helping me to brainstorm and create age-identification strategies. Stephanie, I appreciate your patience with me and my first scheme development project; thank you for pushing me and for weighing in on every agonizing snag with categorization. Jennifer, thank you for donating your precious time and keen eye to create visitor group clips from hours of video. Liz, you are such a fantastic coder, thank you for all the brainpower and energy you put into this project (and now I can finally tell you what it was all about!). Debbie, I admire your ability to hold several disparate schemes in your head, thank you for coding on this dissertation while others coded on yours (and congratulations!). Tommaso, thank you for donating your programmer skills in order to rescue all the video data from StudioCode. Thanks to Veronica, Joyce, and the Geometry Playground team for vetting several aspects of this study and offering their support (in the form of hot sauce and more realistic timelines among other things). I'd also like to extend a general thank you to the Exploratorium and its management for supporting this study and other individual and departmental investigations that foster creativity and understanding.

Finally, my family and friends have been incredibly understanding and helpful throughout my years as a graduate student. A huge thanks to each of them for enduring my ridiculous lack of availability and still providing their support. Jessica, thank you for flying down here to pick up and organize the pieces of my life, for the countless hours hashing out the details of this project, and for giving me energy in so many ways. Kristin, thank you for being there from the very beginning, for providing a place to stay and a car to drive when I came up for my meetings, and for your thoughtful feedback. Wendy, thank you for the innumerable hours spent head down at Ritual and all of the other coffee shops around San Francisco silently motivating me, and for going through it first so that I can rest assured that I am not crazy. Stephanie, thank you for all of the lovely dissertation care packages, remnants of which provide glimpses of your support every day. Thanks to Angi and Tommy for contributing a fresh perspective on this work and for the eight-minute breaks throughout my dissertation days and nights. Kerry, thanks for being willing to set aside your high-powered job at the drop of a hat to help me understand V-lookups. Sarah, thanks for showing up at the very beginning of our friendship to help take photographs for the label, and for providing a lovely retreat for the last few days of writing. Andy, thank you for offering your editorial eye. This list is long, yet I am certain

I've missed some very important contributors; please know that you are not forgotten or unappreciated.

Table of contents

Page

List of tables

List of figures

Designing Exhibits For Gender Equity
Toni Nicole Dancu

## Chapter 1: Introduction

Gender equity in science, math, engineering, and technology (STEM) has been a hard-sought goal over the past half century. Research suggests that improvements have occurred: grade-school achievement in math and science courses is more equitable, and females are now obtaining more than 50% of the degrees in science and engineering (e.g., Coley, 2001; National Science Board, 2008); however the foci of these degrees are more often in social and biological sciences, while the foci for males are more often in engineering, physics, and computer science. While women are now more likely to obtain degrees in the engineering, computer and physics areas of science, they are still earning a smaller percentage of these degrees than are men, are earning a smaller percentage overall of science and engineering master's level degrees, and are still underrepresented in these careers (National Research Council, 2009; National Science Board, 2008). Spatial ability contributes to STEM achievement and gender differences in STEM achievement. Many researchers have shown that spatial reasoning is critical to learning, and obtaining careers, in science, technology, math, and engineering (Battista et al., 1998; Ben-Chaim et al., 1989; Pallrand and Seeber, 1984; Tartre, 1990; Tracy, 1987; Wai, et al., 2009). While recent research has found few areas where sex differences in cognitive abilities exist (Ruble, et al., 2006), the largest and most consistent sex difference in cognitive ability is found in spatial ability, where males outperform females in nearly all aspects (Kimura, 1999; Ruble, et al., 2006).

One of the sources frequently suggested for these observed differences in spatial reasoning may be found in the kinds of informal and play activities in which boys and girls typically engage (Bjorklund & Douglas Brown, 1998; Quaiser-Pohl & Lehmann, 2002; Voyer, Nolan, & Voyer, 2000). Boys tend to play more games that involve construction and gross motor three-dimensional activities (Ruble, et al., 2006). Informal learning environments such as interactive museums provide opportunities to engage in these kinds of experiences and are significantly related to STEM career interests (Hamilton, et al., 1995; Salmi, 2001, 2002). However, girls report significantly fewer visits to science centers (based on National Education Longitudinal Study data from 1988 referenced in Hall & Murphy, 1996; Lee & Burkam, 1996; National Science Foundation, 2003). Further, girls' lower scores on spatial-mechanical reasoning tests have been related to reports of fewer museum visits (Hamilton, et al., 1995).

The conceptual experience of exhibits for girls who do visit science centers may also be notably different from those of boys. Girls often get displaced at exhibits and spend less time at exhibits, leading to fewer experiences (Milgram, 2005; Taylor, 2005, 2006). Girls also have different preferences in exhibit topics, which can lead to lower attraction and time spent at some exhibits (Greenfield, 1995; Kremer & Mullins, 1992). In research regarding naturally occurring family interactions at interactive science exhibits, parents were three times more likely to explain the exhibit's science content to boys than to girls, but equally likely to read the label or explain how to use an exhibit

(Crowley, 2000; Crowley, Callanan, Tenenbaum, & Allen, 2001 p. 258). Crowley and his colleagues at the Children's Discovery Museum in San Jose (Crowley, 2001; Schneider & Cheslock, 2003) found that including a female mascot in the label at each of the exhibits (a cartoon girl, Power Girl) led to significantly more science explanations by parents to their daughters. However, it is not clear whether those increases in explanations reflect the effect of the mascot on the parents directly or indirectly through their daughters' increased interest, or whether this is a transactional effect.

Designing exhibits for gender equity may help to reduce the gender gap in informal science education and have more distal effects on spatial reasoning abilities and female interest and participation in STEM careers (Hamilton, et al., 1995; Salmi, 2001, 2002). In reviewing the literature, and interviewing gender-interested informal learning experts, several approaches to designing female-friendly exhibits have been suggested, but most have not been systematically studied. A review of the literature and best practices in the field has revealed that exhibits designed with female audiences in mind should incorporate features (see Table 2) to achieve design goals in the following three areas:

- encourage social interaction and collaboration;
- connect to social applications and provide context; and
- seek balanced representations of males and females.

This study was part of a larger National Science Foundation (NSF) funded exhibition development and research project, *Geometry Playground,* at the San Francisco Exploratorium. Geometry provides an ideal topic to study exhibits designed with females in mind because of the consistently discrepant nature of its core content, spatial reasoning, in favor of males. Many of the new exhibits for *Geometry Playground* were designed with these features in mind. This research compared girls' engagement and social interactions with their caregivers at two versions of an exhibit, one that did not incorporate any female-friendly design features (Non-Featured), and one that was enhanced to meet the design goals by adding several female-friendly features (Female-Friendly Featured). Boys' engagement and social interactions with their caregivers at both versions of the exhibit were studied to ensure that the female-friendly design features did not introduce any unintended negative effects. The social interaction portion of this study sought to conceptually replicate previous findings regarding the effects of female-friendly design on caregivers' level of explanations to their children (K. Crowley, personal communication, October 22, 2007;Crowley, 2001; Schneider & Cheslock, 2003), and sought to build on those findings by looking more broadly at children's verbal contributions to those social interactions. Video data of visitors at the exhibits was the primary means for determining exhibit impacts on engagement and social interactions for girls and boys.

This dissertation is rooted in a systems approach. The systems science nature of this study is embedded throughout the literature review, including the study's theoretical grounding in Bronfenbrennarian and sociocultural psychological approaches, and the interpretation of prior work leading to several variables of interest in the current study, which search specifically for evidence of transactional effects. Systems models are used to depict and elucidate the study design. Further, the topic of the focal exhibit, mechanical linkages, is a mechanical system, and the development of the Female-Friendly Featured exhibit (via iterative evaluation) addressed the issues related to the systemic nature of an exhibit (that is, small changes to parts of the exhibit lead to major changes in the overall exhibit experience). Finally, the interpretation of the findings from this study draw attention to the systems aspects and their contribution to the field.

This document begins with a literature review, which outlines the importance of the present study by providing evidence of and links between gender disparities in STEM careers, spatial reasoning, and museum experiences. The literature review concludes with the research questions and hypotheses, followed by a chapter detailing the research design and methodology. The remaining two chapters provide the analytical approach along with results, their discussion and a review of the strengths, limitations, and implications of the study.

## Chapter 2: Literature Review

This literature review examines evidence supporting the present study. Initially, the uniqueness of the gender gap in STEM career participation and spatial ability are discussed, followed by a description of those discrepancies. The two most common theoretical explanations for the spatial reasoning gap are overviewed—providing support for the need to create female-friendly science and geometry exhibits. The conceptualization of sex and gender is presented along with a defense of the use of sex as a proxy for gender in the current study. These discussions of gender are followed by a description of a developmental theory that is sensitive to gender norms and helps understand why we might see gender differences: Contextualism. This Contextualist lens is applied to consider and support the study's research design and coding methodology. Finally I review gender gap issues seen in museums, followed by ways to draw from the fields of education, psychology, and museum studies to help museums develop exhibits that address the gender gap in museums, and possibly help mitigate the gap in STEM career achievement and spatial reasoning ability. Appendix A provides a logic model outlining the goals, actions and outcomes addressed in the following literature review.

### The Gender Similarities Hypothesis

As discussed in the introduction, gender differences found years ago are diminishing. Females are receiving more STEM degrees, and performing better in the math and science domains (Shibley Hyde, Fennema, & Lamon, 1990; Voyer, Voyer, & Bryden, 1995). It is important to acknowledge gender similarities in a literature review and research study focused on gender differences and gender equity. In a recent review of

46 meta-analyses, including 124 effect sizes, Janet Shibley Hyde (2005) looked at effect sizes across the lifespan. She examined average effect sizes across a wide-range of research foci, including cognitive variables, communication, social and personality variables, wellbeing, and motor behaviors. Hyde found that the majority (78%) of effect sizes reported were small ($.11 \leq d \leq .35$; 48%) or near zero ($d \leq .10$; 30%). These small effects include average effect sizes of mathematical computation and problem-solving. However, the size of these effects varied by age and test, and at times reached moderate effects (Shibley Hyde, 2008; Shibley Hyde, et al., 1990).

Shibley Hyde's current research program aims to support her Gender Similarities Hypothesis—not in the service of burying actual differences, but to highlight the similarities and areas of gender equity. The Gender Similarities Hypothesis was generated to reduce negative impacts due to over-emphasizing gender differences, such as self-fulfilling prophesies, and lowered parental expectations and encouragement, that can lead to bigger gaps than actually exist (Shibley Hyde, 2005). With Shibley Hyde's concerns in mind, the current project explores STEM career participation and spatial reasoning, areas where gender differences remain.

**Science, Technology, Engineering, and Math (STEM) Careers: The Gender Gap and the Importance of Spatial Reasoning and Interest**

While the discrepancy between males' and females' STEM participation in educational and career pursuits has decreased over the past 20 years, gender differences still remain. For example, females are now obtaining more than 50% of the degrees in science and engineering (e.g., Coley, 2001; National Science Board, 2008). Nevertheless, the foci of females' degrees are more often the social and biological sciences and less often in engineering, physics, and computer science. Over the past few decades, women have become more likely to obtain degrees in the engineering, computer and physics areas of science. However, women are still earning a smaller percentage of these degrees than are men, are earning a smaller percentage overall of science and engineering master's level degrees, and are still underrepresented in these careers (National Research Council, 2009; National Science Board, 2008). From 1999 to 2002, national data indicated that the percentage of women in academic STEM careers actually decreased from 46 to 24 percent (National Research Council, 2009; National Science Foundation, 2002a). Additionally, among employed scientists and engineers, women were more likely than males to be working in educational settings and less likely than males to be employed in business or industry (as cited in National Science Foundation, 2002b). A quote from the Committee on Equal Opportunities in Science and Engineering, as reported to the US Congress in 2000, conveys the importance of this gender gap: "For the United States to remain competitive in a global technological society, it must take serious steps to encourage [women and minorities] to enter [STEM] fields" (p. 41, as cited in National Science Foundation, 2002b). The gender gap in STEM career interests and achievement has been related to girls' interest in related topics and to spatial reasoning ability (Ceci & Williams, 2007; Jaschik, 2007; Wai, et al., 2009; White, 1992).

STEM career pursuits are related to broader interests in science, technology, engineering and math (Lubinski & Persson Benbow, 2006; Wai, et al., 2009), and females may be less interested in these topics (Morgan, Isaac, & Sansone, 2001). For example, Jacobs, Finken, Griffin, and Wright (1998) surveyed over 200 girls and their parents; results indicated that interest in science was more strongly related to science career preferences than the girls' science GPAs, participation in math and science activities, perceptions of friends' support for career choice, and parents' perceptions of students' science ability. Similarly, Morgan et al. (2001) found that college women were less likely to report plans to enter science and math careers (compared to other fields), and these reports were significantly related to perceived interestingness of the career, even after controlling for perceived competence. Hill, Petus and Hedin (1990) administered the Science Career Predictor Scale to 522 middle and high school students; females scored significantly lower on the career interest factor. STEM career interest has been positively related to science museum visits (Hamilton, et al., 1995; Salmi, 2001, 2002). In addition to getting girls interested in STEM topics as an important way to begin reducing the STEM career gap; studies also support the notion that enhancing spatial reasoning abilities may help mitigate the STEM career gap.

Studies relating spatial ability to STEM careers have been ongoing since the 1950s (Super & Bachrach, 1957; Wai, et al., 2009). Many researchers have identified a positive relationship between spatial reasoning ability and STEM achievement and career pursuits (Battista, Clements, Arnoff, & Van Auken Borrow, 1998; Ben-Chaim, Lappan, & Houang, 1985; Pallrand & Seeber, 1984; Tarte, 1990; Tracy, 1987; Wai, et al., 2009). Spatial ability has been identified as one of the personal antecedents necessary for STEM career interests and ambitions. Lubinski and Persson Benbow (2006) analyzed the data from 400,000 youth who participated in an eleven-year longitudinal study and identified spatial ability as critical to STEM educational and occupational achievement. The relationship between spatial ability and career achievement is important because there is also a persistent gender gap in favor of males in spatial ability (discussed in detail below). It is plausible that addressing that spatial reasoning gap could play a substantial role in addressing the gap in STEM career choices.

**Spatial Ability: The Largest and Most Persistent Cognitive Gender Gap**

The exhibit chosen for this study is entitled *Geometry in Motion*, and is part of a larger exhibition entitled *Geometry Playground*. Geometry is a particularly challenging area of mathematics when designing for females because females have historically obtained lower scores on measures of geometric outcomes such as spatial ability and spatial reasoning (Linn & Petersen, 1985; Ruble, et al., 2006; Voyer, et al., 1995). Over the years, a great many researchers have studied gender differences in spatial ability, how those gender differences change by age and by task, and a multitude of explanations for this discrepancy.

In 1985, Linn and Petersen conducted a meta-analysis to address gender differences in spatial ability, with a focus on separating out the types of spatial ability measured. The authors identified three major aspects of spatial ability, and calculated the overall effect sizes of 172 studies using Hedges' branching method:

- <u>Spatial perception</u>: the ability to identify spatial relations with respect to participant's own physical position, in the presence of distracting information. The authors suggest that spatial perception tasks require participants to rely, at least partially, on kinesthetic cues. Measures of spatial perception include the Rod and

Frame Test (Witkin, Dyk, & Faterson, 1962) and the Water Level Test (Piaget & Inhelder, 1967). Linn and Petersen (1985) found significant sex differences in measures of spatial perception, with an average effect size (d) of .44.

- Spatial visualization: the ability to determine a correct solution by manipulating spatial information over several stages. Such tasks require an analytic strategic approach. Tests of spatial visualization include the Embedded Figures Test, the Paper Folding Test, and the Hidden Figures Test. Linn and Peterson found the lowest overall effect size, which was not significant, in this grouping, d = .13. It is important to note that the sample for this grouping was not proven to be drawn from the same population and, according to Hedges' branching method, should not be grouped (Voyer, et al., 1995).
- Mental rotation: the ability to quickly (as measured by reaction time) and correctly mentally rotate 2D and 3D figures (Linn & Petersen, 1985). Tests of mental rotation include the Mental Rotation Test (Shepard & Metzler, 1971; Vandenberg & Kuse, 1978) and the Primary Mental Abilities spatial subtest (Thurstone & Thurstone, 1941). The strongest sex differences were identified in the area of mental rotation, with an overall effect size (d) of .73.

Ten years later, Voyer, Voyer, and Bryden (1995) conducted a follow-up meta-analysis using the same three categories of spatial ability described above. In this analysis, the authors included a greater number of studies (n = 286) and adhered more strictly to Hedges' homogeneity criterion. Overall, the meta-analysis revealed that males performed significantly better than females on measures of spatial ability (d = .37).

However, these findings were qualified by the type of task and the age of the participants. In order to establish that the effect sizes were drawn from the same population, they had to partition the results by participants' ages, and sometimes by specific tasks within each age group. While Linn and Petersen (1985) reported that sex differences in mental rotation and spatial perception exist across the lifespan, Voyer et al. (1995) found a more complex story.

Voyer et al. (1995) found that effect sizes in all three aspects (spatial perception, spatial visualization, and mental rotation) tended to increase with age. Often the age at which sex differences were detected depended on the outcome measure. For example, the youngest age differences were seen at 7 years for the Rod and Frame Test, at 9 years for the Water Level Test, at 10 years for the Primary Mental Abilities Spatial Relations subtest, at 13 years for the Differential Aptitude Test Spatial Relations subtest, and at 14 years for the Embedded Figures Test. Notably, very small sex differences were found in the children's version of the Embedded Figures Test and the Block Design test. Based on the wide variation in these findings, the authors suggest that sex differences in early childhood are "not convincingly established" (p. 261). Given these findings, the current study focused on the interactions of children older than seven years of age (i.e., 8-12).

Findings reported in the most current Handbook of Child Psychology (Ruble, et al., 2006) suggest that the degree to which males outperform females in most aspects of spatial ability varies across domains. The largest sex difference is found in the aspect of mental rotation, especially in three dimensions: d = .56 – 1.00 for adolescents and adults, and about .40 for children. Differences in spatial perception, which requires vertical and horizontal recognition, are about .40, and higher in the specific task of hitting a target with a ball (d = 1.0). The one spatial domain in which females outperform males is in tasks that require memory of spatial location (d = 1.0).

The size of the gender gap also depends on the year of the study. Typically, gender differences are shrinking or at least remaining stable over time. Voyer, Voyer, and Bryden (1995) looked into cohort effects and sex differences on scores. Again, the results varied by task. Some tests revealed a significant negative linear relationship between gender differences on scores and year of birth, that is, diminishing gender differences (including, Identical Blocks Test, Water Levels Test, and Embedded Figures Test). Only the Mental Rotations Test showed a significant positive linear relationship between gender differences on scores and year of birth, or increasing gender differences. The remaining tests showed non-significant trends between sex differences and scores by year of birth. The tests with stable sex differences, in favor of males, were the Primary Mental Abilities Spatial Relations subtest, the Mental Rotations Test, and the Rod and Frame Test.

Generally, it has been established that gender differences exist in the realm of spatial reasoning. The gap between the sexes seems to widen with age, but remains stable or shrinks over the years (with the one exception being the Mental Rotation Test). Gender differences are more prominent in particular measures of spatial ability. The question remains: when we do see gender differences in spatial ability, why are they there? In an effort to explain the sex differences in spatial ability, research has investigated several possible causes. Explanations include a biological divide between males and females, such as hormones or cerebral organization; and differential environmental experiences, such as toys or games that provide spatial skills practice.

**Theories of gender differences: Plausible explanations.** In 1974, Maccoby and Jacklin published the seminal piece, the Psychology of Sex Differences, which concluded that males outperform females on spatial and mathematical tests (as cited in Baenninger & Newcombe, 1995). Research that followed focused initially on biological differences. However, over the years researchers began to look at more socio-cultural explanations as well. Voyer, Voyer, and Bryden (1995) provide a list of topics that have been investigated as possible causes of sex differences in spatial ability. Their list includes: sex hormones, cerebral lateralization, rate of maturation, genetic complement, differing experiences and socialization, differing strategies, and sex role identification (Costa, Terracciano, & McCrae, 2001). While a review of all of the above causes is beyond the scope of this literature review, two common classes of theories, biological and experiential, are outlined below.

***Biological explanations.*** Support for biological theories regarding sex differences in spatial ability is found in studies of hormonal influence and brain lateralization. Burton, Henninger, and Hafetz (2005) based their research on prior findings that larger amounts of hormonal androgens present in females during pre- and post-natal development were related to better scores on measures of spatial ability. Biological research has established that females typically have larger finger length ratios (between second finger and the third, fourth, and fifth fingers, respectively) than males, due to hormonal exposure during and after pre-natal development. Accordingly, Burton, Henninger, and Hafetz found that males with larger ratios (i.e., less male-typical) and females with smaller ratios (i.e., less female-typical) scored higher on mental rotation (measured by the MRT). That is, more developmental androgynous hormone levels during development, assessed with adult finger-length ratios, are related to higher mental rotation scores.

Bryden (1982) provides a good history of hypotheses and findings in regard to gender differences in brain lateralization. In the 70's, two theories began vying to explain sex differences in cognitive functioning through differential lateralization. The first, put forth by Buffery and Gray in 1972, argued that females were more lateralized (certain functions are represented in a single hemisphere rather than in both hemispheres), as demonstrated by their earlier language acquisition. Proponents of this argument suggested that it was beneficial to have language committed to a specific hemisphere, but detrimental to have spatial ability committed to a specific hemisphere. The second theory, put forth by Levy in 1972, held that females were less lateralized, at least in language ability, and that the bilateralization of language processing interfered with the lateralized processing of spatial tasks. In his book on laterality, Bryden (Bryden, 1982) reports that more studies support the latter theory than the former.

In a statistical review of 396 studies on gender differences in lateralization Daniel Voyer (1996) again found support for Levy's theory that males are more lateralized than females. Voyer reviewed articles that used right- or left-field stimuli presentation, based on the assumption that processing occurs on the opposite hemisphere of the brain, and assessed reaction time and correct identification of verbal or non-verbal stimuli. Voyer found that men are more lateralized than women in visual and audio modes of processing regardless of verbal or non-verbal presentation of stimuli. However, the effect sizes are small (d = .076 for the largest overall difference). Voyer suggests that the small effect sizes reflect the inconsistent findings in studies of functional lateralities (see Voyer & Bryden, 1990 for an exemplary study with inconsistent findings). As Voyer concludes, these meta-analytic findings at least partially support the plausibility that cerebral organization (i.e., greater laterality in males) helps to explain sex differences in both verbal and spatial cognitive abilities.

One way to determine whether the gender gap in spatial ability is biological is to look at research across multiple cultures. Unfortunately, very few studies have been conducted outside of Western cultures or across multiple cultures, and those that exist yield conflicting results. Research has identified similar significant gender differences in multiple cultures. Researchers in China found gender differences, favoring males, on mental rotation tests among gifted students in Hong Kong (Baranowski & Delorey, 2007). Researchers investigating menstrual cycle effects on Turkish medical students' mental rotation scores found that males scored significantly higher on the Mental Rotation Test, and that hormone-related effects were small and non-significant (Halpern & Tan, 2001). Mann, Sasanuma, Sakuma, and Masaki (1990) found that, in both Japan and the U.S., high school males scored significantly higher than females on measures of mental rotation, while the opposite was true on the verbal fluency test . Flaherty found significant gender differences on the Mental Rotation Test after matching males and females from Ireland, Japan, and Ecuador in terms of age, education, and social class (2005). All of these results support biological explanations for the gender gap in spatial reasoning.

However, other research has shown that gender differences in spatial ability are not universal. In fact, Flaherty (2005) found that Ecuadorians had significantly lower test scores than the Irish and Japanese participants, and that Ecuadorian males performed as well as Japanese females. Hanna (1990) analyzed data from the International Association for the Evaluation of Educational Achievement (AEEA; http://www.iea.nl/index.html) and found no significant gender differences on geometry scores in Belgium, England, Finland, Hungary, Japan, Scotland, Sweden, or Thailand. AEEA data were collected from over 74,000 13 year-olds from 18 countries. Data included results from five areas of mathematics achievement. Overall gender differences in mathematics achievement were very small across the eighteen countries, and geometry was the only area where these main effects were significant (Hanna, 1990). While half of the countries studied revealed no gender difference in geometry scores, the other half showed significant gender differences in favor of boys. The countries with significant gender differences on geometry were: Canada, France, Hong Kong, Israel, Luxembourg, Netherlands, New Zealand, Nigeria, and the U.S.A. A curriculum analysis revealed that geometry and measurement, the areas with the largest gender differences, were the topics least taught in classes. In a separate study conducted with 19 non-missionized Auca Indians (the entire adult population), females actually outperformed males on spatial representation and construction tasks (Pontius, 1991). However, Pontius includes few details on her methodology. The fact that these gender differences are inconsistent across cultures suggests that while the difference may be partially biological, experience plays a large role in the development of the spatial reasoning gender gap and environmental interventions can enhance spatial abilities (Casey, L., & Pezaris, 1999).

***Experiential explanations.*** Experiential theories of gender differences focus on proximal (direct experiences) and distal (social role expectations and gender modeling) explanations (Costa, et al., 2001). The nature of this research project necessitates a focus on the former; however, the latter is lightly addressed in the *Museum Gender Gap* section below and is theoretically believed to color visitors' behaviors and interactions in their proximal experiences. There are two common ways to study proximal experiential effects on the spatial reasoning gender gap; the first is to correlate past experience with spatial reasoning scores, and the second is to experimentally introduce training and measure the effects of training on males and females (Baenninger & Newcombe, 1989). In a review of the effects of proximal environmental factors on spatial and mathematical skills, Baenninger and Newcombe (1995) report that environmental influence is necessary for the development of spatial and mathematical skills for both genders. Unfortunately, children experience little formal exposure in schools, and spatial ability is more often viewed as innate (Baenninger & Newcombe, 1995; Hamilton, et al., 1995). Consequently, spatial and mathematical ability is underdeveloped for both girls and boys in the U.S. This underdevelopment is compounded for girls because influential informal environmental experiences are more common for boys than for girls (Baenninger and Newcombe, 1995). Providing positive experiences with mathematics and spatial reasoning may improve abilities for both genders and provide equitable access to females.

Baenninger and Newcombe (1989) conducted a meta-analysis of the relationship between spatial activities and scores on tests of spatial ability. The combined correlations ranged from .22 for spatial visualization, .18 for mental rotation, and .13 for spatial perception. The combined correlation between spatial reasoning scores and masculine-typed activities was significant ($r$ = .11), but the combined correlation for feminine-typed activities was not significant at ($r$ = .06). For females, the combined correlation across all spatial activities was significant (r = .09), while the combined correlation for males was not significant. Of note, the authors do not incorporate data to explain whether these relationships arise due to the sex-typing of the activity or the spatial-nature of the activity, nor do they mention any third variables such as subculture, motivation, or socio-economic status.

Voyer, Nolan, and Voyer (2000) assessed the relationship between prior experiences and two measures of spatial ability that typically result in gender differences. The authors asked 291 undergraduate students to report prior experience with several toys and sports, some of which were spatial and others that were not. The Water Levels Test was employed because prior work had shown this measure was more susceptible to environmental influence, and they incorporated the Mental Rotation Test to see if the results differed between the two tests. Voyer et al. found that gender differences were not seen on the Water Levels Test for females who favored spatial *toys* (but still existed in Mental Rotation Test scores). Gender differences existed on both tests regardless of whether *sports* were typed as spatial or not. Boys performed better on Mental Rotation Test and Water Levels Test if they reported playing spatial sports, but girls' performance did not change with types of sports played. This is contrary to the trends Quaiser-Pohl and Lenmann (2002) found; the relationship between sports and Mental Rotation Test scores may depend on which sports are deemed spatial, or different sports may have differential effects on spatial ability.

Quaiser-Pohl and Lenmann (2002) looked specifically at the relationship between undergraduates' Mental Rotation Test scores and self-reported experiences for 183 subjects. They found that females' *technical activity* experiences were positively related to their Mental Rotation Test results, but technical activities were negatively related to males' scores. Similar findings were seen in computer and sports experiences. A positive correlation between *computer experience* and Mental Rotation Test scores was found for females and no relationship for males. A positive but non-significant relationship was identified for *sports activities* and Mental Rotation Test scores for females, but no mentionable relationship for males. The authors conclude that, "compared to males, females' spatial abilities are extremely vulnerable to and thus modifiable through attitudinal and experiential factors" (p. 245). The results from these three studies support the need for enjoyable spatial reasoning activities, especially for females, such as the science museum exhibit featured in this research.

Each of these studies provides encouraging support for the notion that experience can enhance spatial reasoning abilities. However, correlation does not imply causality. It is important to consider other plausible explanations, such as parental influence in regard to toys, sports, and spatial ability (parents interested in spatial activities may introduce children to particular toys and sports, but also be more supportive when it comes to helping with math homework or explaining spatial relations to their children). The direction of the relationship is also unclear in correlational studies: do children who are good at spatial reasoning choose spatial activities, or does experience with spatial activities lead to better ability? One way to explore directionality of experiential effects on spatial reasoning is to test the effects of spatial ability training on spatial reasoning scores.

Environmental influence can also be more direct. For example, Clements, Battista Sarama, and Swaminathan (1997) worked with eight-year-old children from both urban and suburban classrooms to explore the effects of classroom lessons that focused on depth of exploration, developing children's strategy (with a focus on the construction of units and units of units), and providing meaningful problems (i.e., recontextualization of math activities). The curriculum involved a computer game similar to Tetris entitled *Tetrominoes*. Tetrominoes emphasizes ideas about area, deemphasizes speed, and allows replacement, game repeating, and stepping through the game. The curriculum and computer game enhanced students' awareness of geometric motions (slides, flips, and turns), strategies of tetromino placement, and their ability to discuss spatial ideas. The authors found that girls had lower Wheatley Spatial Ability Test scores both pre and post, but showed post-curriculum gains similar to males. Results from this within-subjects study differ from other studies that have found training to reduce the gender gap (see for instance: Ferrini-Mundy, 1987). The authors also found the largest advantage for males lay in the realm of rapid mental rotation, which is counter to some findings that females have a response time advantage in spatial tasks (Voyer & Bryden, 1990); most likely this discrepancy is due to differences in the tasks (WSAT vs. MRT) and ages of participants (eight year-olds cf. undergraduates).

Baenninger and Newcombe (1989) conducted a second meta-analysis of experimental studies that explored the effects of spatial training and gender on spatial abilities. The authors found some interesting combined effect sizes, and differences between the combined effects when broken down by intensity of training. The combined effect size for brief training periods was significant at approximately .53, and significantly smaller than the combined effect size for longer training periods, which was significant at approximately .62. The combined effect size for general training was significant at approximately .50, and significantly smaller than the combined effect size for more specific training, which was significant at .68. The combined effect size for females was larger than the combined effect size for males (e.g., .74 for females and .71 for males in specific training), but not significantly larger—that is, training seems to be equally beneficial for both genders. These studies support the notion that both males' and females' spatial abilities can benefit from direct experience, and that experience plays a role in spatial reasoning.

The varied research programs that identify and aim to explain gender differences in spatial ability are indicative of the multifaceted and complex nature of this discrepancy. Biological and environmental theories are not mutually exclusive, and it is possible that environmental influences modify underlying biological predispositions (Casey, 1996; Casey, et al., 1999; Costa, et al., 2001). The fact that gender differences have been present across history suggests that educators need to find a way to enhance females' experiences with and understanding of geometry. Studies that have looked into environmental influences are particularly supportive of the role informal education can play in providing equitable access to and interest in spatial reasoning tasks. According to many of the studies above, increased access and interest are likely to reduce the gender gap in spatial reasoning abilities and geometric understanding, as well as positively affect the gender gap in STEM career achievement described above. Such studies support the creation of a geometry-focused museum exhibition that can provide visitors (both male and female) a venue for practicing spatial reasoning skills in an enjoyable environment. However, these studies also suggest that creators of such a venue should be particularly cognizant of the experiences of their female audience members, and should work to develop experiences that attract and maintain females' attention, as well as provide a positive overall experience for the females they attract. This research investigated a single geometry exhibit to study whether incorporating female-friendly design features could enhance girls' engagement with the exhibit and provide them with a positive experience on par with boys' experiences.

**Sex and Gender**

While the current study acknowledges the role of biology in the enduring aspects of the gender gap, the focus remains on experiential influences that maintain or mitigate those gender differences. That said, this study still employed the biologically based measure of sex rather than the experientially based measure of gender. For this study, sex is included as an easily observable, albeit rough, approximation of the more influential aspect of gender (which requires more invasive survey methods that are not popular in a setting such as a science center).

In 1979 Rhoda Unger criticized researchers for their focus on sex differences, and therefore, the biological sources of differences between males and females (Kimball, 2007). Unger urged psychological researchers to shift their considerations to gender differences that arise from socio-cultural interactions. Researchers supported the shift in focus and terminology. The field began to distinguish *sex* differences, which refer more often to biologically based states, from *gender* differences, which refer to the psychological features to which people subscribe and more generally associate with the biological states (Deaux, 1985). Gender studies began to consider gender as fluid and complex (Berg & Lie, 1995). However, Unger herself has acknowledged that the distinction is not as important today as it was in the late 70s when researchers were solely considering the biological sources of differences and not the socio-cultural bases (Kimball, 2007).

Today, researchers acknowledge the fact that sex and gender are not dichotomous, but that the two overlap substantially. As Deaux (1985) asserts in her review, the fact that meta-analytic discrepancies are often small and differences change over time should help to dispel the belief that sex differences represent a bi-modal distribution; there is a lot of overlap in most male and female distributions (Kimbal, 2007). Thus, gender is more accurately understood as a continuum, from maleness to femaleness. In the context of the current research, using sex as an approximation of gender provides a conservative estimate of effects for girls and boys because the more accurate gender affiliations have been averaged over in this dichotomous variable.

**Developmental Theory: Contextualism, Vygotsky, and Socioculturalism**

This research draws on a developmental perspective that can help understand gender related behaviors within the broader system. The developmental theory known as Contextualism maintains that the social contexts a person participates in, particularly with more skilled adults, exert significant influence as learners actively construct knowledge, and that previous knowledge, language, perceptions, and interactions all influence subsequent interests, learning and meaning making (Miller, 1993). That is, the child is not seen in isolation, or simply in an activity, but as an active part of a larger whole. The focus of Contextualist study is the child-in-activity-in-context and "the path from the object to child and from child to object passes though another person" (Vygotsky, 1978, p. 30). With a systems oriented perspective, Contextualists recognize that the child is shaped by his or her social-cultural-historical context and that these contexts are inseparable (Miller, 1993; Rogoff, 1990; Vygotsky, 1978). For example, females at museums bring their active internalizations and interpretations of gender roles and norms to their experience. As educators, we cannot separate the current state of females, their gender role affiliations, and their interests from the politically correct notion that these differences are socially created and should be mitigated to keep from further purporting gender differences. Put simply, as both educators and researchers we have access to girls where they are within a culture, not where we hope they will or think they should be.

According to Bronfenbrenner and Morris (1998) there are many levels within a child's context. Of interest to this study are the more immediate *microsystem* and the more overarching *macrosystem*. The microsystem involves the activities and social interactions that are directly experienced by the child, such as the exhibit activities and the verbal and non-verbal interactions with their caregivers at the exhibits. In the field of Systems Science, for this project the microsystem would be deemed the *system*, with the adult, the child and the exhibit as the *subsystems* (Lendaris, 1986). The macrosystem on the other hand, refers to the patterns of interactions and activities that emerge within the broader social context, such as culture or subculture; the social-cultural-historical context that shapes a child, and thus, a child's experiences (Miller, 1993). For example, within the United States, the macrosystem provides particular belief systems, values, and opportunity structures that guide the social exchanges between caregivers and children. This macro- or *suprasystem*, consciously or unconsciously, influences children's goals, risks, and experiences allowed or encouraged by caregivers (and others, such as teachers); which may differ based on the gender of the child (or the social class) (Lendaris, 1986; Miller, 1993). As can be seen in these examples, the influence of the more distal levels of context, such as culture, are often experienced and influenced by the child through the more proximal levels of context, such as the parent interactions.

Vygotsky viewed learning as the process, not the product—learning stimulates internal developmental processes which operate in a social environment containing more skilled others (Vygotsky, 1978). That is, what the child actually does when engaged in an activity with another person is much more important than correct or incorrect responses or actions (Miller, 1993). Eventually these processes are automated and internalized and the child has advanced developmentally (Vygotsky, 1978). This view is further developed in Vygotsky's belief that instruction should occur at the child's potential level of ability rather than the child's actual level. Such instruction advances the child's development rather than maintaining it.

Vygotsky's most cited theoretical concept is the zone of proximal development. Vygotsky purported that a child could reach a certain degree of understanding or ability alone (Vygotsky, 1978). Yet, when an adult or more skilled child scaffolds—prompts, encourages, and assists—the child, that child can reach their full potential in understanding or ability (Vygotsky, 1978). Full potential is believed to already exist within the child but beyond their immediate grasp without scaffolding (Vygotsky, 1978). The difference between the child's actual ability and the level of competency possible when scaffolding is present is the zone of proximal development (Vygotsky, 1978). It is important to recognize that the child is not a passive recipient, but actively participates in explorations that engage their social partners (Miller, 1993). The child's personal attributes can, therefore, encourage or discourage reactions from their caregivers that facilitate or thwart development (Miller, 1993). This study sought to determine whether girls elicited more scaffolding behaviors from caregivers at the more female-friendly exhibit.

In most literature, scaffolding is provided by an adult or more experienced child, however, many believe Vygotsky's scaffolding referred to "…any situation in which some activity is leading children beyond their current level of functioning" (Miller, 1993, p. 384). In the context of this research, it may be the parent or the exhibit activities that will scaffold the child's level of understanding. A caregiver can scaffold a child through "prompts, clues, modeling, explanation, leading questions, discussion, joint participation, encouragement, control of the child's attention, and so on" (Miller, 1993, p. 379). The social aspect of learning is inherent to development, and Vygotsky argued that learning naturally occurs when children engage in conversations with more knowledgeable others. A well-designed hands-on exhibit allows the child, or the adult and child, to question, investigate, and manipulate the properties at a higher level to allow the child to reach a greater level of understanding. "Science centers are envisioned to entice learners to go beyond their present knowledge and to construct a newer, larger vista of scientific thinking" (Ramey-Gassert, 1997, p. 436). An exhibit can also provide the caregiver an entry point for scaffolding the child. The present project observed video interactions of children and their caregivers at exhibits to determine the degree to which caregivers scaffold their children and the ways that children question and talk about the exhibit activities.

Vygotsky's sociocultural theory also helps explain why we may see gender-particular behaviors by understanding that the culture in which one lives permeates all aspects of that person's life. Culture is considered a medium through which experiences are perceived, interpreted, understood, and enacted (Miller, 1993). This is important in understanding how children raised in a particular culture, with particular gender norms, are expected to respond and interact within a culture that emphasizes different values, communication, and skills (all of which Vygotsky deemed "tools"). Cultural views of female-appropriate roles and interests are prominent for most females. As early as two years of age, children are able to identify their sex and begin categorizing the world using gender as a guide (Ruble, et al., 2006). Gender-norm activation can influence how a learner perceives, attends to, interacts with, and discusses an exhibit. If an educational venue such as a museum can approach a topic in a manner that helps females to sidestep and reinterpret such cultural norms, then that venue can help enhance interest and understanding for its female visitors.

A child's culture will provide them with psychological tools, which will influence both thought and behavior in the learning process (Miller, 1993). For instance, language is thought to be the most powerful psychological tool provided by a culture, shaping a child's thought, attention, perception, behavior, speech, and goals (Kozulin, 1986; Miller, 1993; Vygotsky, 1978). Children use their speech along with perceptual cues (e.g., visual and tactile) to explore phenomenon and solve problems. Leinhardt and Crowley (1998) define learning as conversational elaboration that is part of the process and product of exhibit exploration. This study focused on the verbal tools available to children to understand their processes at the exhibits.

Contextualist theory provides support for the experiential theories explaining gender differences in girls' and boys' spatial reasoning abilities. This developmental theory also broadens these explanations by looking at the cultural context. Using Contextualist theory, we gain insight into why girls' and boys' interactions with museum exhibits may differ when the exhibits are designed with different socio-cultural goals in mind. Finally, Contextualist theory provides a strong argument for observing verbal interactions between children and their caregivers in the context of each exhibit, and offers ample guidance for interpreting those proximal processes.

**The Museum Gender Gap and Design-based Solutions**

The research reviewed up to this point suggests that informal experiences play a key role in reducing the spatial reasoning and STEM career gaps. However, some researchers have identified discrepancies in science museums' male and female attendance and visitor experiences. Girls have reported significantly fewer visits to science centers (based on National Education Longitudinal Study data from 1988 referenced in Hall & Murphy, 1996; National Science Foundation, 2003). Worse, reports of fewer museum visits have been related to girls' lower scores on spatial-mechanical reasoning tests (Hamilton, et al., 1995), even after controlling for possible spurious correlations including ethnicity and socio-economic status. In re-analyzing the Philadelphia/Camden Informal Science Education Collaborative (PISEC) data, Borun found that "parents are less likely to bring their daughters to science museums than their sons" (Borun, 1999, p.12). More boys than girls were seen in the Science Museum [M = 425 (58%): F = 302], Natural Science Museum [216 (56%): 170], and aquarium [239 (53%): 215], while no gender differences were recorded in zoo attendance [437 (50%): 443]. Pilot data from the Oregon Museum of Science and Industry revealed significantly more male youth than female youth (Benne, personal communication, October 12, 2007). Other research indicates that while boys and girls report a similar quantity of informal science experiences outside of school (listed as any related life experience), they differ on type (Erickson & Farkas, 1991). Of interest, boys' experiences are more applicable to science achievement. It is unclear whether females are less often at the museums because of marketing, parental socialization, or because the exhibits are not representative of their interests and learning/play needs, and therefore do not keep them coming back. This study addresses the latter explanation.

Once at the museum, girls have differential experiences with the conceptual aspect of the exhibits. Researchers and evaluators have found that, at times, girls have different preferences in exhibit topics, which can lead to lower attraction or holding time at some exhibits (Greenfield, 1995; Kremer & Mullins, 1992). Diamond (1994) reviewed twelve studies of family behavior at science museums, and found that that eight of the studies revealed discrepant experiences for males and females. She found research suggesting that boys interact with exhibits more independently than do girls, and that in the museum context, mothers and fathers engage their sons and daughters differently.

Crowley and his colleagues (Crowley, 2000; Crowley, et al., 2001) analyzed video to study 298 naturally occurring family interactions at 18 interactive science exhibits in a children's museum. Parents were three times more likely to explain the science to boys than to girls, but equally likely to read the label or explain how to use an exhibit to boys and girls (Crowley, 2000; Crowley, et al., 2001 p. 258). Borun's PISEC data do not support Crowley et al.'s findings; however, the PISEC data were measured differently because they were not initially collected to address this issue (Borun, 1999). Crowley and his colleagues (2000, 2001) then investigated children's contributions to this finding by analyzing the number of children's questions in the ten-second and 60-second periods prior to parental explanations, and found no significant relationship. Crowley and his colleagues later found that including a female mascot at each of the exhibits (a cartoon girl in each label) lead to significantly more science explanations by parents to their daughters (Crowley, 2001; Schneider & Cheslock, 2003). It is not clear whether increases in explanations reflect the effect of the mascot on the parents directly or indirectly through their daughters' increased interest. It is possible that parents' explanations were based on their children's level of interest in the exhibit concepts. Crowley did explore the children's' questions preceding parental explanations and found no effects. The current research aimed to look more closely at the reciprocal effects of caregiver and child interactions. Specifically, this study investigated the number of Statements and Questions posed by children and the types of Explanations offered by adults, to determine whether these interactions differ when the exhibit includes female-friendly design features.

Women who bring their children to science centers also have and model a different experience than men who bring their children to science centers. Taylor (2005) found that when observing families, the boys were usually first to use the exhibit, followed by the girls, then the dad, and the mom often gave up her turn so that the family could move on. In exhibit evaluation, anecdotal evidence from interviewers suggests that women commonly defer to their male partners (personally, I've noticed that this occurs even if the male accompanying her is her young son), asking him to answer for her, or telling the evaluator that he knows better than she (J. Gutwill, personal communication, October 7, 2007). These interactional dynamics affect the woman's experience at the science center, and may send an unintentional message to their daughters who are primed to watch their female role model for gender cues (Martin & Ruble, 2004; Taylor, 2005). While beyond the scope of the current study, research is needed to understand how the female-friendly design features (described below) affect female caregivers' interactions with and at science exhibits. Museum educators have spent years creating enticing educational exhibits for everyone. Over the past fifteen years, the field has learned more and more about ways to bring girls in and avoid female-unfriendly exhibits.

**A possible solution: Designing female-friendly exhibits.** In reviewing the literature, and interviewing gender-interested informal learning[1] experts, several approaches to designing gender equitable exhibits have been suggested. While most reports and colleagues' experiences (based on their own observations and interpretations of the literature) resonated with one another, the majority of approaches have not yet been systematically studied. This research, evaluation, and anecdotal review has revealed that exhibits designed with female audiences in mind should incorporate features to achieve the design goals in the following three areas:

- encourage social interaction and collaboration;
- connect to social applications and provide context; and
- seek balanced representations of males and females.

Below, each design goal, in these three areas, is discussed in detail along with specific design features for addressing those goals, supportive studies from the fields of education, museum studies, and psychology, and followed by suggestions, examples, and lessons learned from museum studies colleagues (i.e., what worked and what did not work). An evaluation of the strength of evidence for each goal is provided based on the

[1] Informal learning (also referred to as free-choice learning) is used to define learning by choice that occurs outside of school; this is the type of learning often occurs, and is studied, in museums, zoos, aquaria, after-school programs and other similar ventures (National Research Council, 2009).

quality of science behind the studies (e.g., depth of exploration, breadth of participants, and type of research design) and the number of convergent studies. Whenever information is specific to *girls* or *women*, those terms are included; otherwise, the review refers more generally to *females*. Table 2 summarizes the suggested design goals and features.

***Encouraging collaboration.*** Creating exhibits that encourage collaboration by allowing for shared goals in the exhibit experience can help developers reach three design goals: fostering social interaction, avoiding speed and competition, and enhancing girls' experiences when boys are involved in the activity. Each goal is detailed below along with specific design features.

*1. Foster social interaction*: Exhibits are more successful with female audiences if they are built to elicit group or cooperative learning. Females tend to work together on exhibits and support one another in the process (Diamond, 1994; Finn, personal communication, July 17, 2007; Froschl, Sprung, Archer, & Fancsali, 2003; Maher, 2005; Milgram, 2005; National Science Foundation, 2003; Rosser, 1991; Taylor, 2005).

<u>Strength of evidence</u>: Moderate to High

While the majority of these reports are conglomerations of evidence that provide too little detail to evaluate the quality of the science supporting their conclusions, a few studies provide a deeper level of detail. Taylor's research (2005, 2006) incorporated simple counts of visitor behavior using video data from over 400 visitor groups; however, he doesn't report any statistical or descriptive information to back his findings of collaborative female and independent male groups. Educational researchers employing experimental designs found that females are more likely than males to develop mathematical and spatial skills in groups than when alone (Friedman, 1995 as cited in Clements, et al., 1997; Phelps & Damon, 1989). A small quasi-experimental study employing both quantitative and qualitative methods found that girls in grades 1, 3, and 5 performed better on the Water Level Test when paired with a friend; however, boys performed worse because their talk became off topic (Kutnick & Kington, 2005). This finding suggests that researchers need to determine the effects of more social exhibits on boys' on- and off-task conversations. The current study checked for any adverse effects of the features on boys' engagement and social interactions.

<u>Applied design features include:</u>

- Creating activities that require two people (two parts to be done simultaneously) (Koke, personal communication, September 13, 2007).
- Employing two-person benches, which provide cues that the exhibit is for more than one person (Koke, personal communication, September 13, 2007). However, it is important to consider the trade-off between benches and whole-family access, wheelchair access, face-to-face interactions, and flexibility in design (Benne, personal communication, October 12, 2007). Two-person benches may be hinged to swing toward and away from the exhibit, offering access to larger groups or visitors using wheelchairs.
- Employing the family friendly suggestions identified during the Philadelphia/Camden Informal Science Education Collaborative (PISEC) research (Borun, personal communication, September 14, 2007):

  - multi-user;
  - multi-sided—visitors can cluster around the exhibit;
  - comfortably accessible to adults and children;
  - relevant—provides cognitive links to visitor experiences and knowledge, that is, connect the exhibit to their lives (see also, National Science Foundation, 2003);
  - multi-outcome—the activity is complex, which engenders group discussion.
- Encouraging talk; collaborative talk can be as important as collaborative exploration. Museum labels can model the types of questions that provoke discussion at exhibits, as can questions posed during programs and workshops (McCreedy, personal communication, October 24, 2007).
- McCreedy, based on experiences with girl/adult collaboration programs, suggests that an activity's social component should be inherent or necessary, rather than arbitrary (personal communication, October 24, 2007).

<u>Successful designs include:</u>

- Edwin Schlossberg is a developer who emphasizes social interaction (see http://www.esidesign.com) (Koke, personal communication, September 13, 2007). A good example is the Bells in Church Steeple exhibit at the Pope John Paul II Cultural Center. An overhead speaker is activated when a visitor enters the exhibit space and explains that six visitors are needed to play a song; each visitor pulls a bell rope when their indicator lights up. With fewer than six people, the song is incomplete. Visitors have been known to run around the museum asking strangers to join them.

- The Space Odyssey exhibition in Denver was designed with females in mind. Women preferred exhibits that placed the family in team roles (e.g., astronaut team roles) (Koke, personal communication, September 13, 2007).
- In approaching the topic of sound vibration activity stations during an outreach program, the Franklin Institute (FI) found that cup-and-string telephones foster collaboration. However, the same topic broached with a tuning fork and water did not require two participants, and arbitrary collaboration did not work well for participants (McCreedy, personal communication, October 24, 2007).

Unsuccessful designs include:

- None noted.

2. *Avoid speed- or competition-based activities:* Females frequently report negative experiences with competitive interactions often found in math and science (Rosser, 1991; Taylor, 2005). However, it seems girls do enjoy group competition, but do not wish to be pitted against each other in individual competition (Koke, personal communication, September 13, 2007).

Strength of evidence: Low to Moderate

Lessons can be incorporated from the gaming community where, as in math, girls' use of and interest in computers declines dramatically after the age of 13 (National Science Foundation, 2003). After many attempts to design games for girls, three sites have garnered success: Sims, Whyville, and Purple Moon. The successful components of these games are further discussed below, but for now it is important to note that all three games are non-competitive, and feature a variety of experiences and possible outcomes (Jenkins, 2001).

Educational research provides a broader view into competitive or cooperative learning environments. Peterson and Fennema (1985) conducted a study to identify whether there were classroom activities differentially related to boys' and girls' math achievement. The authors conducted a pre-test and a six month follow-up post-test (using NAEP items), along with observations of time engaged (or not) in several activities for focal students from 36 $4^{th}$ grade classrooms in rural Midwestern towns. Girls' post-test achievement (controlling for pre-test) showed the strongest negative relationship to engagement in competitive learning activities (even more than off-task behavior). Not surprisingly, girls' engagement in cooperative learning activities was significantly and positively related to math achievement for both genders. It is important to note, however, that boys' results were opposite those of girls. Interestingly, girls' and boys' engagement in activities that required competition amongst groups (i.e., competitive and collaborative) was significantly and positively related to their math achievement. These results suggest that when there are gender discrepancies in educational activities, removing or incorporating cooperative or competitive elements (depending on the direction of the gap) could help reduce the gap, and whenever possible, considering group competition could be the best way to productively engage both males and females.

Applied design features include:

- Building for collaboration and cooperation (see above), which thwarts inclinations to respond with speed and competition at exhibits (Borun, personal communication, September 14, 2007).
- Allow for multiple experiences and outcomes, rather than a single "better" route and end-point (Jenkins, 2001).
- When competitive, consider group competition (Peterson & Fennema, 1985).

Successful designs include:

- None noted.

Unsuccessful designs include:

- None noted.

*3. Enhance girls' experiences when boys are involved in an activity:* In boy-girl interactions (compared to girl-girl interactions), girls generally are afforded less access to the experience, take on fewer leardership roles and are less willing to actively participate (Finn, personal communication, July 17, 2007; National Science Foundation, 2003; Rosser, 1991; Taylor, 2005, 2006).

Strength of evidence: Low

Two different research studies compared all-girl to mixed-gender summer camps; both researchers found that boys dominated the activities, leaving girls to be the notetakers (National Science Foundation, 2003; Rosser, 1991; Taylor, 2006). Yet in all-girls situations, the girls made sure that everyone had a chance to take on each role

(Taylor, 2006). Taylor also conducted interviews with girls at these science camps. He found that many of the girls reported that it was easier for them to speak up and to learn from the all-girls camps because boys were not disrupting them . He provides the following illustrative quote from a girl in one of the mixed-gender camps "we needed to band together to protect ourselves from the boys (p. 195)." Taylor's work was part of a dissertation at the University of Washington, but his pubications do not provide sufficient detail to independently evaluate the methodology employed. Another researcher found that girls in a science center were less likely to use computers, in part because boys pressured girl users to leave, while girls who approached an occupied computer simply moved on (Greenfield, 1995). Greenfield's data were based on systematic observations of 614 children with their caregivers; however, it should be noted that only 34 of those observed were at computer exhibits. Taylor conducted another observational study of sibling groups and found that the boys often displaced their sisters at exhibits (Taylor, 2002). When boys take over, or girls defer to boys, girls get less hands-on time, leading to fewer experiences and less comfort with the appropriate tools, technology, and machines (Milgram, 2005).

<u>Applied design features include:</u>

- Creating exhibits that require two people to actively participate.
- Employing two-person benches, which discourage boys from displacing the girls and help girls create an alliance against anyone who might pressure them to move (Koke, personal communication, September 13, 2007).
- Including an observational aspect allows girls to get comfortable with the activity

prior to approaching it, thus giving them more confidence when using the activity (Rosser, 1991).

Successful designs include:

- None noted.

Unsuccessful designs include:

- None noted.

***Making connections.*** Creating exhibits that are grounded in meaningful context, can help provide entrée for females and offer ways to connect to the an exhibits' content. Developers can make these meaningful connections by connecting the exhibit to social or community applications, or providing context surrounding the exhibit concepts. Each of these design goals is detailed below along with specific design features.

*4. Connect to social or community applications:* Females express a greater interest in learning math and science when the topics include practical applications such as solving social problems, improving/connecting to the lives of people and animals, or exploring community/environmental concerns (D. J. Ford, Brickhouse, Lottero-Perdue, & Kittleson, 2006; Froschl, et al., 2003; Jones, Howe, & Rua, 2000; Kekelis, et al., 2005; Koke, 2005; Maher, 2005; Rosser, 1991; Taylor, 2005). For example, McCreedy (2005) found that enlisting women to educate girls in science led to a more sustained commitment to science teaching and learning for the women.

Strength of evidence: Moderate to High

Several of the museum-based studies do not provide enough detail about their methodology to accurately evaluate their conclusions. Of note, Kekelis et al. (2005) held focus groups with youth from the surrounding neighborhoods to identify the needs of young females, and Ford et al. (2006) created surveys for 45 girls and their parents to determine their reading access and preferences. Anecdotally, evaluators have noticed that women often look for links to relevance in their daily lives, asking questions like: 'How does this apply to me?' or 'What does that have to do with everyday concerns?' (Koke, personal communication, September 13, 2007). The separate areas in which males and females receive STEM degrees also reflect females' preferences for social and community application (and provide the strongest data-driven support): females more often receive degrees in biology and medicine, while males more often receive degrees in engineering, computer science, and physics (National Science Board, 2008; Spelke, 2005). Baranowski and Delorey (2007) conducted focus groups, interviews, and a large survey to identify boys' and girls' engineering topic interests; several of the girls' favored topics aligned with the focus on community application (see suggested topics below). Finally, the analysis of the US responses (totaling 437 sixth graders) from an international survey exploring areas of interest in science revealed that males had significantly more interest than girls in 20 areas of science, while girls had significantly more interest than boys in only six areas (Jones, et al., 2000). While there were a few exceptions, the majority of these areas tended toward boys expressing more interest in the physical sciences and girls more interest in the biological sciences (see also, Baker & Leary, 1995). Specifically, girls reported higher interest in learning about healthy eating, animal communications and AIDS (among others, see Use Their Language and Aesthetics section below).

Applied topics include:

- Environmental science: pollution, conservation (Hill, et al., 1990), building alternative-fueled cars (Baranowski & Delorey, 2007), or ecology in the community park (National Science Foundation, 2003).
- Forensic science (Baranowski & Delorey, 2007).
- Medical and veterinary science (Froschl, et al., 2003; Hill, et al., 1990).
- Engineering: designing/building machines that allow sight for the blind (Baranowski & Delorey, 2007).

Successful designs include:

- The Oregon Museum of Science and Industry's (OMSI) *Moneyville* exhibition featured a stock market game that was initially unpopular with girls—evaluators could not even entice girls to participate. Developers made changes with great success in attracting and holding girls' interest in the game. One of the most notable changes tied each company's work to the community (Benne, personal communication, October 12, 2007). Another important change included a more feminine Bear and a more masculine Bull as newscasters.
- The *Hope* exhibit at the Pope John Paul II Cultural Center asks, "what is your hope?" Visitors can type or record their responses on video. Then, they may share their answers or place them in a time capsule (Koke, personal communication, September 13, 2007).

- The three Web game sites that are most successful with female audiences (Sims, Whyville, and Purple Moon) share social and community attributes: all three have social goals, community-focused activities, and realistic and relatable characters (Jenkins, 2001).
- The *New Amazing Machines* exhibition at the Franklin Institute connected the machines to everyday, recognizable objects that were not gender-linked (Borun, personal communication, September 14, 2007).
- *Listen* at the Exploratorium provides personal/social connections by offering video clips of people throughout the exhibit (Borun, personal communication, September 14, 2007).

<u>Unsuccessful designs include:</u>

- None noted.

*5. Provide context:* Girls and women often wish to know the context surrounding concepts and phenomena (Borun, personal communication, September 14, 2007; Froschl et al., 2003; Milgram, 2005), as well as how and why they are doing something prior to engaging in the activity (Rosser, 1991). Context attracts and orients girls to the exhibit, and provides a reason for them to consider the topic important or meaningful to them. Educators often employ storytelling to engage girls by providing a narrative for science or math concepts (Casey et al., 2008; Casey, Erkut, Ceder, & Mercer Young, 2008). The current study explored the effects of adding female-friendly design features on girls' attraction and time spent at an exhibit.

Strength of evidence: Moderate to High

A small survey study regarding elementary students' reading preferences found that when asked about different types of science stories, girls preferred informational narratives (a scientific version of historical narrative), followed by experiment examples (e.g., try this), then fiction, and least preferred information-only science books (D. J. Ford, et al., 2006). Similarly, Comparative Studies Professor Jenkins at MIT, based on his and others' applications of previous theory and research to case studies, has concluded that computer games and Websites popular with females (such as Purple Moon, Sims, and Whyville) incorporate storylines throughout the activity (Cassell & Jenkins, 2000; Jenkins, 2001; Subrahmanyam & Greenfield, 2000).

Beth Casey and her colleagues (2008) report on two studies they conducted to explore the use of storytelling to help children learn geometry skills, with particular attention paid to gender effects. The first study employed a pre-post skills transfer task to compare scores for children who had received geometry lessons embedded in a story to a control group of children who did not receive geometry lessons. Boys' scores improved independent of the lessons/control, but girl's scores only improved when learning geometry lessons embedded within a story. The second study took place in a high-poverty community and assessed pre-post skills transfer for children who received typical geometry lessons and children who received the same geometry lessons embedded in a story. In the second study, neither boys nor girls showed improvement in the geometry-only condition, and both boys' and girls' scores increased when the lessons were embedded in a story; however only girls' increases were significant.

Applied design features include:

- Answering the following questions (Milgram, 2005; Rosser, 1991):
    - What does it come from?
    - What does it contribute to?
    - What does it connect with?
    - How will it be used?
- Making connections to other experiments (Milgram, 2005; Rosser, 1991).
- Making connections to other topics (e.g., how does math relate to sociology, health, economics, chemistry?) (Rosser, 1991).
- Providing a theme or back-story (Borun, personal communication, September 14, 2007).
- Increasing the observational aspect of the activity—this eases girls in, allowing them to become comfortable enough to explore less-familiar tools and technology (Rosser, 1991).
- McCreedy cautions that museums must take care to avoid compromising the math and science content (personal communication, October 24, 2007).

Successful designs include:

- In the science-based *Discovery Days* offered as part of the Girls at the Center program at the Franklin Institute, parents and children work together to build an amusement park structure that draws on their experiences and capabilities. This type of activity introduces key concepts (e.g., marshmallow and toothpicks to demonstrate the strength of a triangle), and offers a scenario with clear roles for the adult and child. This context also allows for creativity and offers infrastructure for collaborative talk among visitor groups (McCreedy, personal communication, October 24, 2007).

- All Franklin Institute exhibitions provide narratives and back-stories (Borun, personal communication, September 14, 2007).
- *Listen* at the Exploratorium provides themes within the exhibition (Borun, personal communication, September 14, 2007).

Unsuccessful designs include:

- None noted.

***Balanced gender representation.*** Providing balanced representations of girls' and boys' interests, language, aesthetics, and imagery within and across exhibits can send a strong message about the intended audience to visitors, and provide an entrée into exhibit topics for a variety of visitors. Developers, writers, and graphic designers can work toward balanced representation by emphasizing cross-gender skills and preferences, using language and aesthetics that appeal to girls as well as boys, and highlighting female role models and users. Each of these goals is detailed below along with specific design features.

*6. Emphasize cross-gender skills and preferences:* Many museum professionals and researchers report gender preferences for certain topics or exhibits (see Table 1 for specific examples). The purpose of exploring gender-typical exhibits is to find exhibits that are attractive to both genders and to balance out the more masculine-typed exhibits with more feminine-typed exhibits. Museums need to emphasize cross-gender skill use, which we may be able to do by making male-associated exhibits more attractive to girls and vice-versa (Diamond, 1994). It is also important to provide female-typed exhibits because the terminology and tools are more comfortable and the topics more immediately accessible to females.

Strength of evidence: Moderate

There is strong evidence to support the idea that girls and boys have different interests in areas of science (see *Connect to Social Applications and Provide Context* above). There is also strong evidence that girls and boys have different experiences when it comes to science (Jones, et al., 2000; see also the Experiential Explanations for the spatial reasoning gender gap in the literature review above). For example, Jones, Howe and Rua (2000) analyzed the responses of 437 diverse US sixth graders to a survey developed for an international study of children's experiences outside of school (among other variables). The authors found that significantly more boys reported experiences with batteries, electric toys, electronics, fuses, microscopes, rifles, pulleys, and tools such as saws, wheelbarrows, car jacks, and axes; while significantly more girls reported experiences with sewing, knitting, weaving, clothes making, planting seeds, bird watching, stargazing and bread making. The multitude of museum-based studies presented in Table 1, while most were single exhibition or institution studies[2], together provide ample evidence that developing exhibits with those interests and experiences in mind can lead to more balanced attraction and holding time at museum exhibits.

[2] Of note, the Technopolis Gender Experiment (Verheyden, 2003) exhibition electronically tracked and compared over 54,000 visitors by gender. Their results revealed minimal gender differences in performance, but several large differences in attraction (i.e., exhibit use).

Suggested topics include:

Table 1

*Topics and exhibits that are interesting to females, males, or both*

| **Topics and exhibits that worked especially well for attracting and holding *females*:** | **Topics and exhibits that worked especially well for attracting and holding *males*:** | **Topics and exhibits that worked especially well for attracting and holding *males and females*:** |
|---|---|---|
| Life science exhibits (Greenfield, 1995). | Physical Science exhibits (Greenfield, (1995). | Designing machines that allow sight for blind people (Baranowski & Delorey, 2007). |
| The chemistry of food and nutrition (Anderson, Zhang, Chatterjee, Robin, & Aldrich, 2005). | Engineering exhibits (Anderson, et al., 2005). | Building alternative-fueled cars (Baranowski & Delorey, 2007). |
| The technology behind a hair dryer (Milgram, 2005). | Electricity exhibits (Verheyden, 2003). | Exhibits about photography (Baenninger & Newcombe, 1995). |
| Solving crimes with DNA evidence (Baranowski & Delorey, 2007). | Space Exploration (Baranowski & Delorey, 2007). | Tennis and ping-pong (Baenninger & Newcombe, 1995). |
| Exhibits about people: Recognizing Faces, Identifying Body Parts as being male or female, and Make A Baby and find out what gender it is (Verheyden, 2003). | The science of sport (Anderson, et al., 2005). | Making a design or picture at the Computer Clubhouse (Gallagher & Michalchik, 2007). |
| Designing a playground for children with disabilities (Milgram, 2005). | Video Game design (Baranowski & Delorey, 2007). | |
| A game with increasingly difficult mental arithmetic tasks (Verheyden, 2003). | Spatial insight exhibits (Verheyden, 2003). | |
| Robots who are involved in performance art (e.g., ballet), or are characterized as animals (Cavallo et al., 2004; Milgram, 2005) | Robots (Anderson, et al., 2005) and programming dinosaur robots (Verheyden, 2003). | |
| Electronic Jewelry Workshop (Sylvan, 2005). | Computers (Greenfield, 1995). | |
| Drawing pictures of a phenomenon (Koke, personal communication, September 13, 2007). | Making a video or animation (Gallagher & Michalchik, 2007). | |
| Animal Lab (Kremer, & Mullins,1992) | Water Jets (Kremer & Mullins,1992). | |
| Face Paints (Kremer, & M ullins,1992). | | |
| Puzzles (Greenfield, 1995) | | |
| Writing a newsletter, article or story (Gallagher & Michalchik, 2007). | | |

Applied design features and approaches include:

- Envisioning the end goal: "to design exhibits that enhance girls' interest in and understanding of science, while not discouraging boys" (Taylor, 2002, p. 7).
- Broadening the topic by incorporating several real-world examples with which everyone has experience (Benne, personal communication, October 12, 2007).
- Seeking problems and examples from more traditionally female-associated fields to provide a balance from the more traditionally male-associated fields; for example, home-economics and nursing (Rosser, 1991).
- Considering ways to turn topics that are male gender-role stereotyped to female stereotyped topics (mixing concrete→mixing cookie batter, building model airplanes→completing a dress pattern) (Rosser, 1991). However, McCreedy cautions that assuming relevance by gender can also backfire—offering multiple entrance points and diverse examples is best (personal communication, May 13, 2008).
- Talk to males and females about possible topics, try a card sort to identify interest across a variety of topics (OMSI Evaluation & Visitor Studies Department, 2008).

Successful designs include:

- OMSI focuses on removing girl-unfriendly exhibits or aspects of exhibits rather than building girl-friendly exhibits (Benne, personal communication, October 12, 2007). As a standard prototyping process, every formative evaluation looks for possible gender differences and addresses any imbalance early on.

- OMSI's Tech Hall incorporates a variety of everyday technological examples, from toilets to teddy bears to computers.

Unsuccessful designs include:

- None noted.

*7. Use their language and aesthetics:* Language is very important when teaching math to girls (National Science Foundation, 2003). It is recommended that text contain language and metaphors that reflect both boys' and girls' experiences (Baranowski & Delorey, 2007). Design aesthetics can also play a role in engaging men and women differently (Moss, Gunn, & Helloer, 2006; Moss, Gunn, & Kubacki, 2007); it seems important to be consciously representative in language and design aspects of exhibits.

Strength of evidence: Low to Moderate

Although many authors throughout the NSF 2003 report emphasize the importance of using females' language, they provide too few details about their methods and study deigns to allow for an evaluation of the quality of the evidence supporting their conclusions. The web design research conducted by Moss and her colleagues provides the strongest evidence for considering female language and aesthetics in development processes. Moss, Gunn, and Heller (Moss, Gunn, & Heller, 2006) began by investigating gender differences in website designers' use of language, visuals, and navigation approaches. The authors applied a previously developed design rating system to 60 randomly selected personal websites (30 male-designed and 30 female-designed), and found that 13 of 23 design approaches were significantly different depending on gender of designer.

Female designers included significantly more:

- Abbreviations;
- Informal language;
- Self-denigration language;
- Links to other subject sites;
- Females and males in photos;
- Colors of font other than black/blue;
- Number of colors represented in the font; and
- Rounded lines in font.

Male designers included significantly more:

- Expert language;
- Formal typography;
- Horizontal lines (the appearance of a horizontal line in the layout);
- Formal images; and
- Use of crests.

The authors later collaborated with analysts in France and Poland and found similar differences in design approaches (Lake, 2005). A follow-up study was conducted to determine whether males and females actually prefer their own sex's design approach (what Moss terms Mirroring). Moss, Gunn, and Kubacki (2007) began by each rating 60 randomly selected sites, based on the 13 criteria above, to ensure accuracy of ratings. The authors then asked 64 students to rate seven sites on a scale of 1-20. Females gave significantly higher ratings to the female designed sites (which featured the female design approaches described above), and males gave significantly higher ratings to the male designed sites (which featured the male design approaches described above). These findings suggest that when trying to design with both males and females in mind, designers should take care to represent language and aesthetics of both genders.

<u>Applied design features include:</u>

- Balancing the use of words like *master*, *command* or *tackle*, with words like *connect*, *choose* or *embrace (National Science Foundation, 2003)*.
- Using gentler terminology such as *seems wrong* rather than *wrong (Mintz, 2007)*.
- According to an MIT comparative studies expert, incorporating female voices in the design and content may be one of the quickest ways to tap interests and language (Jenkins, 2001). It seems important to have mixed-gender teams involved in the development process.
- Considering whimsical and aesthetically pleasing designs. Girl-focused program developers from MIT, Wellesley College, and the University of Colorado have suggested that aesthetics help to connect science projects to more lived-in, everyday environments and are an important aspect in appealing to students (Resnick, Berg, & Eisenberg, 2000). The authors discuss the use of whimsy and aesthetics in the Cabaret Mechanical Theatre (http://www.cabaret.co.uk/) and in marble mazes that include bumpers, ramps, and bells. Additionally, a large study of US respondents (described above; Jones, 2000) identified that girls had a strong interest in learning about the science of color and of rainbows (among other things).

<u>Successful designs include:</u>

- Sims, Whyville, and Purple Moon were created by teams that included female designers with authority. In the case of Sim City, which appeals to both genders, the design was not intended to appeal to girls, but simply achieved this task because there were many highly ranked female game designers on the team:

> "…The decisions they made came out of a context where there were more female designers and more highly ranked female designers than I have seen at any other mainstream game studio. In such a context, even if there is no conscious goal of expanding the female market, the unconscious decisions made by men and women working together is likely to produce a product that is very different from one where the intuitive decisions were made by an all or predominantly male team of designers. Not surprisingly, then, the Sims has proven to be highly successful in attracting female players while at the same time, the product has expanded the range of play experiences available to boys."(Jenkins, 2001).

- Moss's work suggests that including female design approaches could more strongly attract and engage females (Moss, Gunn, & Kubacki, 2007).

Unsuccessful designs include:

- None noted.

*8. Highlight female role models and users:* Young women often hold misconceptions about the lives of women with careers in science; providing pictures and stories about women in science can help to challenge this misconception (Diamond, 1994; Froschl, et al., 2003; Hill, et al., 1990; Koke, 2005; National Science Foundation, 2003).

Strength of evidence: Low

Most of these sources are compilations of multiple findings, but few provide enough detail to ascertain the quality of the science behind these claims. Those that did provide enough information have conducted a small number of focus groups or card sorts that, while helping to identify the impact of role-models, did not engage enough representative participants to broadly generalize. Taylor (2006) conducted 12 in-depth interviews and found that many of the women he interviewed that were interested in science spoke of having “almost all male teachers and very few female role models (p. 195),” while some stayed in the research sciences, some were encouraged to move into the marketing or educational aspects of science. In working with the image of engineering for males and females, the National Academy of Engineering analyzed responses to 12 youth’s responses to trading cards (Baranowski & Delorey, 2007). They found that girls tended to pick images that pictured female engineers, while boys were more likely to pick images of objects. The Power Girl study conducted by Crowley and his colleagues found that including an image of a female mascot in the exhibit labels significantly and positively impacted the number of science explanations parents provided to their young daughters (K. Crowley, personal communication, October 22, 2007; Crowley, 2001; Schneider & Cheslock, 2003). While these results are promising, they have not yet been published and thus, a review of the methodological quality is not possible.

<u>Applied design features include:</u>

- Providing examples based on the work of female scientists, mathematicians and designers (Benne, personal communication, October 12, 2007; Rosser, 1991; McCreedy, personal communication, October 24, 2007).
- Girls seem to more easily relate to women close to their age and more like them (McCreedy, personal communication, May 13, 2008). Including examples of their hobbies and interests also seems to make science more accessible.

- Including images of females in the labels, based on Power Girl findings (K. Crowley, personal communication, October 22, 2007; Crowley, 2001; Schneider & Cheslock, 2003).

Successful designs include:

- A study by Girl Scouts USA found that girls are more influenced by 18-25 year-olds who do interesting, science/math-related things, than by more established individuals (Borun, personal communication, September 14, 2007; McCreedy, personal communication, May 13, 2008).

Unsuccessful designs include:

- The *Air Show* at the Franklin Institute incorporated female stories and photos into labels throughout the room (very close to the exhibits), but a summative evaluation tracked the use of these labels and found that nobody read them (Borun, personal communication, September 14, 2007).
- As mentioned above, an image sort revealed that girls were more likely to choose images of female engineers, while males were more likely to choose objects (Baranowski & Delorey, 2007). However, boys and girls disliked pictures of individuals that were sitting or standing at a desk.

Table 2

*Suggested design features to achieve female-friendly design goals*

| | # | DESIGN GOALS | SPECIFIC DESIGN FEATURES | USE |
|---|---|---|---|---|
| BUILDING FOR COLLABORATION | 1 | **Encourage social interaction and collaboration** | a) two simultaneous parts of activity<br>b) two-person benches<br>c) multi-outcome<br>d) connect to their everyday lives<br>e) multi-user/sided | Exhibit and Label |
| | 2 | **Avoid speed- or competition-based activities** | a) two simultaneous (collaborative) parts of activity<br>b) everyone wins<br>c) multi-outcome (rather than succeed/fail)<br>d) group/team competition is great | Exhibit |
| | 3 | **Enhance girls' experiences when boys are involved in the activity** | a) two simultaneous parts of activity<br>b) two-person benches<br>c) incorporate an observational aspect | Exhibit |
| MAKING CONNECTIONS | 4 | **Connect to social or community applications** | a) practical applications:<br>-solving social problems,<br>-improving/connecting to the lives of people and animals,<br>-exploring community and environmental concerns<br>b) offer links to relevance in daily life | Exhibit and Label |
| | 5 | **Provide the context surrounding concepts** | a) where does it come from?<br>b) what does it contribute to?<br>c) what does it connect to?<br>d) how will it be used?<br>e) make connections to other exhibits<br>f) provide a story<br>g) make interdisciplinary connections<br>h) incorporate an observational aspect | Exhibit and Label |
| BALANCED REPRESENTATIONS | 6 | **Emphasize cross-gender skills and preferences** | a) talk to girls and boys to identify interest across a variety of topics<br>b) seek problems/examples from traditionally female associated fields, too (home-ec, nursing, life-sciences)<br>c) consider offering male- and female-typed aspects (mixing concrete→cookie batter) | Exhibit |
| | 7 | **Use their language and aesthetics** | a) balance out the use of words like *master*, *command* or *tackle*, with words like *connect*, *choose* or *embrace*<br>b) "seems wrong" or "may be wrong" rather than "wrong"<br>c) include female designers on the team<br>d) informal language<br>e) colored and rounded fonts | Exhibit and Label |
| | 8 | **Highlight female role models and users** | a) provide pictures of young female scientists/mathematicians<br>b) provide their stories, too<br>c) provide pictures of female users | Label |

The literature reviewed for this research project reveals geometry to be a particularly important area for providing positive informal experiences for females. In identifying ways to reach girls through such informal geometry experiences, we find multiple design goals that should engage females and encourage equitable social interactions at exhibits, along with many suggestions for design features that may reach those goals. However, while the design goals have stronger research and practitioner support, the design features are mostly lacking in any systematic investigative support. The present study seeks to determine the effects of incorporating several of the aforementioned female-friendly design features.

**Purpose of the Present Study**

This study compares a geometry exhibit that is typically low on female-friendly design features (Non-Featured version), to the same geometry exhibit when enhanced with multiple female-friendly design features (Female-Friendly Featured version). A between-subjects research design was employed to determine whether incorporating these design features enhances the exhibit experience for girls, and to explore whether such changes affect boys' exhibit experiences. The focus of this research is on creating exhibits that are equitable for girls. Currently, many science exhibits are inequitable, showing favorable engagement and social interactions for boys. Therefore, the main research questions and hypotheses focus on the effects of the exhibit conditions on girls' Engagement and Social Interactions. However, to ensure that the exhibit changes do not have negative effects on boys' Engagement and Social Interactions, each hypothesis for girls is followed up by an expected null result for boys. While it is unusual to hypothesize null results, they can be very informative. Cook suggests that a researcher should not only test the outcomes expected to be affected by the treatment, but also gain discriminant validity for their results by testing outcomes that are *not* expected to be affected by the treatment (West, Biesanz, & Pitts, 2000). To that end, I hypothesized an additional null result, expecting that the changes to the exhibit would affect the number of Meaningful Explanations parents provide their girls, but not the number of Direction/Procedure Explanations parents provide their girls. Figure 1 provides a system overview of the variables important to the study (A. Ford, 1999).

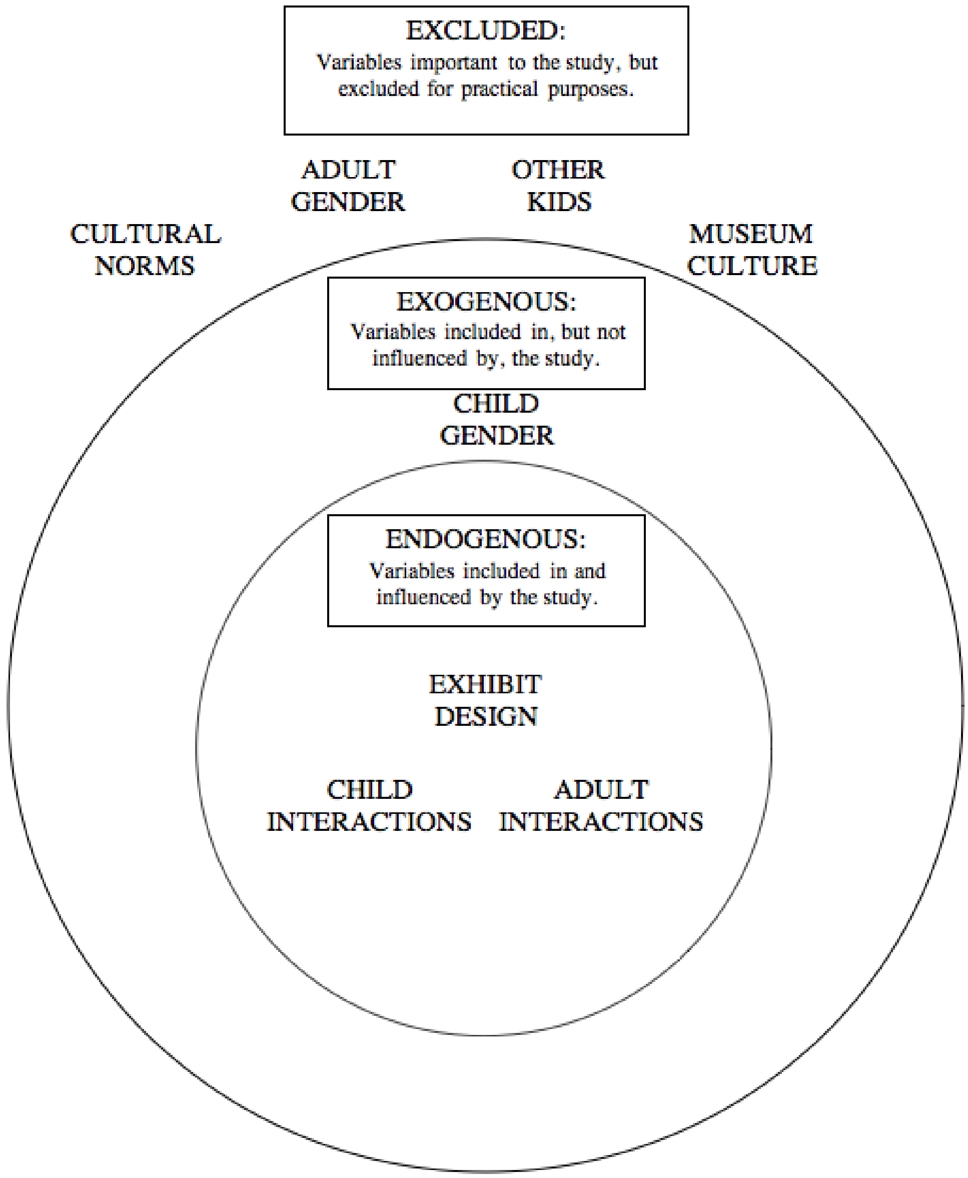



Figure 1. Systems bull's-eye diagram of variables important to the study.

Video and audio data of visitors at the exhibit were the primary means for determining the impacts of incorporating female-friendly features on girls' (and boys') Engagement (Attraction and Duration) and Social Interactions. The specific research questions and hypotheses are detailed below.

**Research Questions and Hypotheses**

1. Does girls' Engagement with the exhibit (Attraction and Duration) depend on whether Female-Friendly design Features are incorporated?

   *Hypothesis 1a*. Girls were expected to be more likely to stop at the exhibit once the design features were incorporated.

   *Hypothesis 1b*. Boys were expected to be equally likely to stop at the exhibit regardless of the design features.

   *Hypothesis 2a*. Girls were expected to spend significantly more time at the exhibit once the design features were incorporated.

   *Hypothesis 2b*. Boys were expected to spend similar amounts of time at the exhibit regardless of the design features.

2. Does the quality of girls' Social Interactions with caregivers at the exhibit depend on whether Female-Friendly design Features are incorporated?

   *Hypothesis 3a*. Girls were expected to utter significantly more Statements at the exhibit once the design features were incorporated.

   *Hypothesis 3b*. Boys were expected to utter a similar number of Statements at the exhibit regardless of the design features.

*Hypothesis 4a*. Girls were expected to ask significantly more *Q*uestions at the exhibit once the design features were incorporated.

*Hypothesis 4b*. Boys were expected to ask a similar number of *Q*uestions at the exhibit regardless of the design features.

3. Does the quality of caregivers' Social Interactions with girls at the exhibit depend on whether Female-Friendly design Features are incorporated?

*Hypothesis 5a*. Caregivers were expected to provide girls significantly more *Meaningful Explanations* once the design features were incorporated.

*Hypothesis 5b*. Caregivers were expected to provide girls a similar number of *Directions/Procedure Explanations* at the exhibit regardless of the design features.

*Hypothesis 5c*. Caregivers were expected to provide boys a similar number of *Meaningful Explanations* at the exhibit regardless of the design features.

*Hypothesis 5d*. Caregivers were expected to provide boys a similar number of *Directions/Procedure Explanations* at the exhibit regardless of the design features.

*Hypothesis 6a*. Caregivers were expected to ask more Questions of their girls once the design features were incorporated.

*Hypothesis 6b*. Caregivers were expected to ask a similar number of Questions of their boys regardless of the design features.

Figure 2 provides a flow diagram of the hypothesized relationships among study variables for girls.

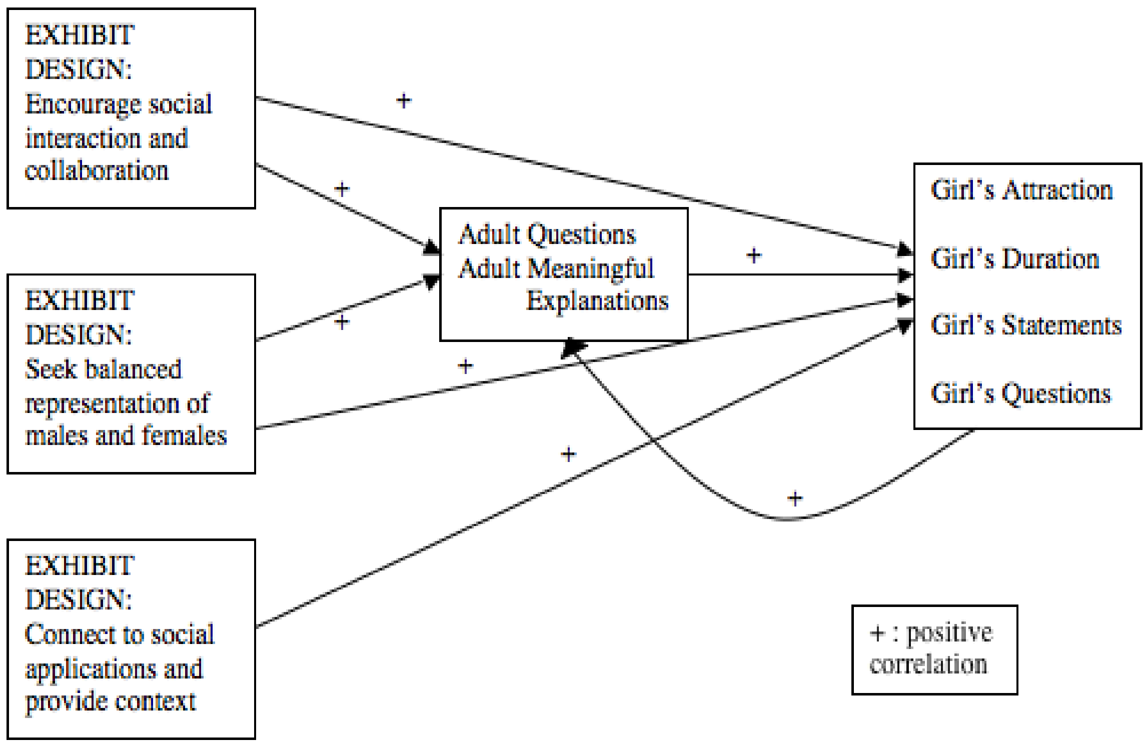



Figure 2. Flow diagram depicting hypothesized results for girls.

# Chapter 3: Methodology

This research investigated the behaviors of science museum visitors in two versions of a geometry exhibit, one of which was specifically designed to be engaging for girls, and to encourage equitable social interactions with their caregivers. Engagement and Social Interactions with a caregiver at each version of the exhibit were video and audio taped. To ensure no negative effects of including female-friendly design features occurred for boys, their Engagement and Social Interactions were also explored. This chapter describes the participants, exhibits, experimental design, procedures, and methods that shaped this research and aided in determining the impact of the female-friendly design features.

## Setting

The Exploratorium is a museum of science, art, and human perception founded in 1969 by physicist Frank Oppenheimer. The Exploratorium's mission is to create a culture of learning through innovative environments, programs, and tools that help people nurture their curiosity about the world around them. The Geometry Playground project is a National Science Foundation funded exhibition development and research project that aims to design gender-equitable exhibits that foster spatial reasoning.

The specific geometry exhibit for this study, described below, was set-up in a special section of the Exploratorium, the Sound Abatement area. This area is open on three sides, carpeted and has special wall coverings to reduce the amount of ambient noise making it better for video and audio recordings. Stanchions were placed around the area in order to create two entrances. Signs informing visitors that research was taking place and that they would be recorded in the area were placed next to the entrances. See Figure 3 for a drawing of the Sound Abatement area.


Figure 3. The sound-abatement area of the Exploratorium.

**The Exhibit**

*Geometry in Motion* is an exhibit about mechanical linkages and linkage construction. A linkage is “A system of interconnected machine elements, such as rods, springs, and pivots, used to transmit power or motion” (The American Heritage® Dictionary of the English Language, retrieved September 07, 2008). The *Geometry in Motion* exhibit consists of rods, wheels, spacers and magnetic connectors or pivots (see Figures 4 and 5). Visitors can connect any of the pieces to each other or to the table. The goal of the exhibit is to encourage people to explore and build their own ideas, but specific machine examples are provided in the labels. It is worth noting that this exhibit was identified on the Computing for Sustainability blog as one of the few exhibits at the

Exploratorium to provide visitors experience with complex systems (http://computingforsustainability.wordpress.com/2008/07/30/). However, the construction and machine aspects of this exhibit made it a challenging exhibit to make female-friendly.


Figure 4. Original *Geometry in Motion* exhibit.

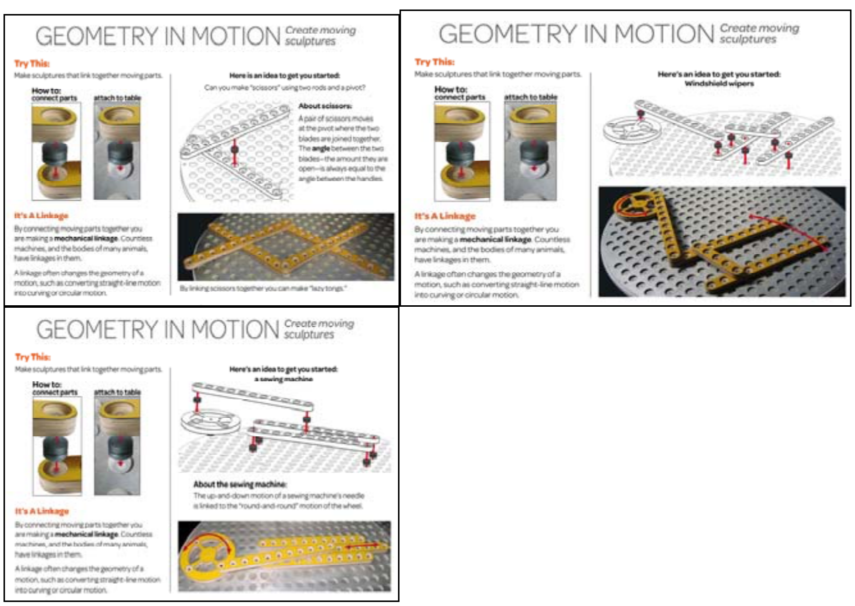



Figure 5. Original *Geometry in Motion* labels.

**The Non-Featured condition.** The original *Geometry in Motion* exhibit was created with several female-friendly features in mind. I worked with the exhibit developer to re-configure the original version to be more representative of typical museum exhibits and to remove the female-friendly features that were incorporated. This Non-Featured version is a single-station, single user exhibit (see Figure 6). The label depicts a single machine, windshield wipers, as an example. The label does not refer to social connections such as joints in humans or animals as real-world examples of linkages (see Figure 7). The aesthetic of the exhibit (yellow rods with silver table-top) remained in its originally developed state.

Figure 6. Non-Featured *Geometry in Motion* exhibit.

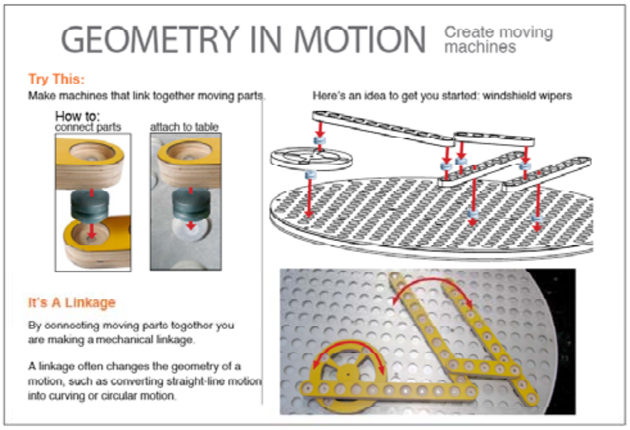



Figure 7. Non-Featured *Geometry in Motion* label.

**The Female-Friendly Featured condition.** The female-friendly version of the exhibit was developed to incorporate several of the features aimed at addressing the female-friendly design goals that fall under the flowing thee areas: encourage social interaction and collaboration, connect to social applications and provide context, and seek balanced representations of males and females. Development of the female-friendly condition began with a select advisory team who helped determine which features made the most sense for this particular exhibit, and which changes would be most representative of each of the goals and features. The team consisted of a female Project Director who has worked on labels and exhibits at several science and children's museums across the nation, a male and a female exhibit developer, a female evaluation and label specialist, a male senior researcher, a female artist, and a female field trip manager—all are employees at the Exploratorium and all have a special interest in reducing the museum gender gap.

***Considerations for reaching female-friendly design goals***. Several possible changes to the exhibit features were considered. The full list of considerations, based on advisory team meetings, is included in Appendix B. Decisions for the implemented changes were based on anticipated impact, cost and developer time, as well as formative evaluation with visitors on the floor whenever possible(Scriven, 1991). In an effort to encourage social interaction and collaboration (including ownership of the area), the female-friendly exhibit became multi-station (two identical work stations at the exhibit) and multi-sided (approachable from more than one direction), and enabled visitors to build across stations collaboratively. The exhibit itself was tilted and the labels were made bigger and more colorful, providing visitors a better understanding of and comfort with the activity prior to use, if so desired. The label also suggested that visitors create something less concrete than machines—*sculptures*— where there is no success or failure inherent in the activity. In order to make connections, the label included several pictures of real-world linkages (a chair, a lamp, train wheels, and the prosthetic knee on a horse). Exhibit components and labeling were designed to encourage people to build sculptures, which provides visitors an opportunity to create their own narrative or storyline at the exhibit, as did the new hammer and bell components. This version included more aesthetically appealing whimsical components using multiple colors for new pieces and more relatable features, such as hands, eyes, and feet. In order to balance masculine and feminine representation, relational vocabulary was incorporated in the label, and photos in the label were neutral, male, and female in their associations (e.g., a folding chair, train wheels, and a horse with a prosthetic leg). The label also depicted a drawing of a female using the exhibit components, thus providing a subtle invitation to girls and hopefully encouraging their caregivers to provide explanations to their girls while at the exhibit (as with Power Girl; K. Crowley, personal communication, October 22, 2007; Crowley, 2001). Finally, there was considerable consensus among females on the advisory team that the exhibit was too messy to be approachable; the female-friendly version included a tilted table and a much easier storage system for the parts. See Figures 8 and 9.

***Color selection***. A formative evaluation was used to determine which four of the seven Geometry Playground exhibition colors would be used for the new exhibit pieces. Ten girls between the ages of 8 and 12 were approached separately and asked to help us decide which two colors to use in addition to the yellow we already had for a new exhibit. The girls were handed one of the original yellow rods (exhibit components), and then the seven pantone chips were laid out and the girls were asked to pick one additional color. Once they had selected a top choice, they were asked to pick one more color to add to the yellow rod in addition to their initial color choice. Their choices were compiled by adding up the number of times a color was mentioned, regardless of order. Green, pink, orange and light blue were ranked the highest, and were therefore included in the exhibit components. Purple, white, and dark blue had the lowest rankings.

***Iterative formative evaluation.*** Any change to an exhibit can significantly change visitors' ability to use the exhibit; I conducted several iterative evaluations to modify the exhibit and its label until visitors were able to use the new version successfully (see Appendix C for a full list of evaluation iterations and changes). I employed the Rapid Iterative Testing and Evaluation methodology typical in Human-Computer Interactions and museum exhibit development evaluations (Dick & Carey, 1996; Medlock, 2005; Medlock, Wixon, Terrano, Romero, & Fulton, 2002). This method helps identify major issues quickly and recognizes that you only need to test iterations with 3-5 people to uncover primary problems. Usually, an initial iteration is tried with visitors, and areas where visitors have difficulty using the exhibit are quickly changed in response to the nature of their struggle. For example, if a visitor approached the exhibit and connected components to one another, but never connected components to the table, we would adjust our communication regarding the connection to the table. This process therefore requires several in-the-moment changes. Iterative development continued for three months and was completed when most visitors were able to use the exhibit successfully.

Figure 8. Female-Friendly Featured *Geometry in Motion* exhibit.

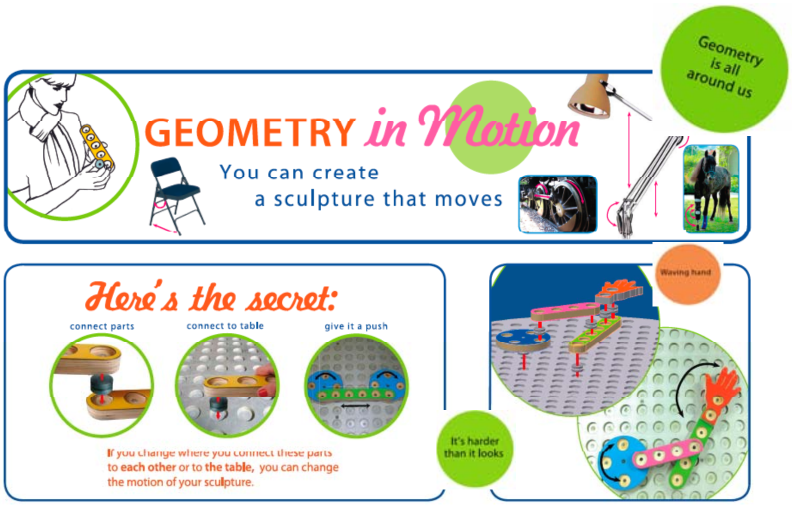



Figure 9. Female-Friendly Featured *Geometry in Motion* label.

**Experimental Design**

This quasi-experimental design consists of two exhibit conditions, one low on female-friendly features and one high on female-friendly features. The exhibit, *Geometry in Motion,* was adjusted for each condition. The between-subjects design implies that each child visited only one version of the exhibit. Video and audio data were collected at each version of the exhibit. Qualitative data were coded according to a coding scheme and analyzed quantitatively.

**Participants**

Participants were visitors to the Exploratorium, one of the original science centers, located in San Francisco, California. The Exploratorium does not currently have estimates of visitor demographics; however, admissions records reveal that 51% of visitors are adults, and 43% of visitors receive free or discounted admission. Visitors may be museum members, general Bay Area visitors (50% of visitorship), or out-of-town visitors (Exploratorium, 2008).

A power analysis, using independent samples t test formulas (Howell, 2002), was conducted to determine the minimum sample size per treatment group. Power is the probability of correctly rejecting the null hypothesis. A power analysis can be used to determine an appropriate sample size by holding the power constant and using an estimated effect size for the study. Convention and practicality support the use of power = .80 in determining sample size. One of the best ways to estimate the effect size for a study is to determine the effect sizes in prior research similar to one's planned research.

For the purposes of this study, I collected a sample size of a approximately 78 for each gender in each exhibit condition based on an estimated effect size of .45. This sample size applied to the smallest condition, engaged dyads; therefore, a larger sample size was collected for Duration data because this data set includes the children with caregivers and children who use the exhibit alone. Based on the results of Shibly Hyde's recent gender similarities findings (2005), a small effect size of .25 or .30 should be used estimate the appropriate sample size in each analysis. However, in comparing laboratory and field studies, Deaux (1985) suggests that small effects in the laboratory translate to egregious effects in the real world, due to the multiply determined nature of behaviors in the real world. Laboratory studies tend to focus on capability or what participants can do, whereas field studies tend to focus on what participants actually do. Accordingly, research in the museum studies has led to fairly large effect sizes.

Several studies of similar topic or approach have reported approximate effect sizes ranging from .48 to 1.77 (Crowley, et al., 2001; Humphrey & Gutwill, 2005; Wilde & Urhahne, 2008). For example, Crowley and his colleagues (2001), found a moderate effect of child's gender (d = ~.48) on parental explanations at science exhibits. In a research study of the differences between open-ended Active Prolonged Engagement (APE) exhibits, and single-outcome Planned Discovery (PD) exhibits, Gutwill (2005) found that visitors asked three times as many questions at APE exhibits (d = ~1.77); spent significantly more time at APE exhibits (d = ~1.64); and were more likely to ask explanation-type questions (d = ~.49). Based on these studies, a medium effect size of at least .48 should be used for this study. Table 3 provides several examples of possible effect sizes and corresponding sample sizes per exhibit condition. The final effect size used to calculate the sample size in this study was .45. The final sample size included 345 children for the Attraction data, 260 children alone and 323 children with their caregivers for the Duration data, and 323[3] children with their caregivers for the Social Interaction data. Therefore, for the Social Interaction data for example, the sample size was 78 males at the Non-Featured version, 88 males at the Female-Friendly Featured version, 74 females at the Non-Featured version, and 83 females at the Female-Friendly Featured version.

[3] An additional 19 videos were assigned to coders, but were not coded and dropped from the analyses due to audio (9) or due to confusion regarding the speaker or target of the Adult Informative Talk (10) for more than 20% of the video. Eleven Female-Friendly Featured exhibit videos were dropped (5 boys and 6 girls), and 8 Non-Featured exhibit videos were dropped (2 boys, 6 girls).

Table 3

*Sample size needed based on varying effect sizes with Power set at .80.*

| d = effect size | n = sample size (per condition) |
|---|---|
| .15 | **697** |
| .30 | **175** |
| .35 | **128** |
| .40 | **98** |
| .45 | **78** |
| .48 | **69** |

Power cannot be calculated for hypotheses of no differences. This research is focused on whether female-friendly design features added to the exhibit enhance girls' Engagement and Social Interactions at the exhibit. Subsidiary to the main research goal is the desire to ensure that the female-friendly design features have no unexpected adverse effects on boys' Engagement or Social Interactions at the exhibit. Given that the latter is an expected null effect (and therefore, incalculable), power was estimated using the effect sizes expected for girls.

**Implied Consent**

The Exploratorium's umbrella protocol for conducting research with human subjects requires that visitors included in identifiable video data be informed of the recording and its purpose, and also be able to decline participation. The Exploratorium typically informs visitors of research recordings by posting a sign at the front of the museum, cordoning off the area under study and placing a sign at each entrance explaining the recording and its purpose, and leaving the microphones and video camera in plain sight. Adding small signs informing visitors of the video recording to each exhibit and stanchion, raised visitors' understanding that the area was being taped for research purposes from 75% to 99% (Gutwill, 2003). The current research study incorporated all of the above signage methods to inform visitors of the research and video recording in the cordoned-off area.

Video and audio recording occurred on weekend days during the school year (used for pilot and training data) and Tuesday through Sunday throughout the summer months (used for the research analyses), ensuring that adults accompanied minors in the area requiring implied consent. All video recording was terminated no later than 4:00 pm. The early termination allows visitors who do not want to participate in the research, but wish to experience the exhibits inside of the stanchions, to return during their museum visit.

**Procedures**

Video and audio recordings were collected in the Sound Abatement area of the Exploratorium. As described in the implied consent section above, this area of the museum was cordoned off to create two entrances with adjacent signs informing visitors about the research. Two additional exhibits were included in the Sound Abatement area, Making Waves and Hyperbolic Slot. Making Waves is a series of 20 identical magnetic pendulums dangling side-by-side from a stable spine; the movement of any one pendulum impacts the entire system of pendulums in interesting ways. The Hyperbolic Slot exhibit allows visitors to surprisingly push a straight rod through a curved cutout in a plastic plane because a line is a 2D slice of a 3D hyperboloid (see Figures 10 and 11).

These exhibits were kept in the same location in the Sound Abatement area for the entirety of the study.

Figure 10. Making Waves.

Figure 11. Hyperbolic Slot.

The video data was collected Tuesday through Sunday for one version of *Geometry in Motion*, and on matched days the following week for the second version, over six summertime weeks. The number of days needed to collect data depended on the number of visitors selected each day, in order to reach the target sample size (a variation of quota sampling; Visser, Krosnick, & Lavrakas, 2000). Days were divided into five hour-long units, from 11:00 am to 4:00 pm. For each matched pair of data collection days, each hour was assigned a randomized number. The hour assigned the highest random number was also used for *Attraction* data for that matched pair of days. The entire five hours of each day's video was used to select visitors for *Duration* and S*ocial Interaction* data.

This research used height as a developmental approximation of age, rather than using the more intrusive procedure of asking visitor's ages in real-time, in an effort to reduce the impact on visitors and to increase the number of visitor groups attainable on each day. Using pediatric charts I was able to identify heights that had high inclusion for the ages of interest and low inclusion for those who were too young or too old. These pediatric charts also indicated ages with less discrepancy in height by gender, allowing me to choose the ages/heights that were least likely to confound the results of the study. The heights chosen for this study target children ages 7-13, with the following makeup: 47" and up includes some seven year-olds (38%) and most 8 year-olds (92%) (but, only 18% 6 year-olds and a negligible number of 5 year-olds); 62" and under includes some 13 year-olds (50%) and most 12 year-olds (75-80%) (and only 25% of 14 year-olds and all 11 year-olds). If a visitor was in the height range, but clearly older than 12, they were not included in the data; that is, observational judgment was sometimes necessary even with the height approximation. The overall effect was a sample of children mostly between the ages of 8 and 12, but with some 7 and 13 year olds. The fact that age was not a variable of interest in this study, and that exhibit exploration and social interactions do not vary dramatically within this age range, makes it likely that this approximation method does not impact on the results of the study.

A fish-eye lens was placed on the video camera in order to capture (in a single image) visitors' entry into the Sound Abatement area and exhibit use. To ensure that accurate heights were captured while using the image-altering fish-eye lens, master height videos were created at the beginning of data collection (and any time the camera in the area was moved) along with an accompanying overlay image. A master height video required that two data collectors stand at either entrance to the Sound Abatement area (and at another location further in from the entrance), and hold a level across the entrance at 47" and again at 62" for about ten seconds (see Figure 12). The master height videos were then used to create overlay images (see Figure 13), where electronic lines were created to mimic the level (see Figure 14). These overlay images were then placed over every video for collecting data on Attraction and for visitor group selection (see Figure 15).

Figure 12. Master video of level at entrance.

Figure 13. Create overlay image on Master.

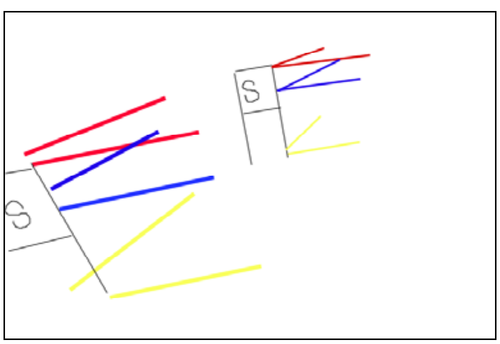



Figure 14. Overlay image.

Figure 15. Apply overlay image to video.

Focal children (and their caregivers) for Duration and Social Interaction data were selected from video data, a selection process employed successfully by Humphrey and Gutwill (2005) and accepted in the informal learning research methodology. A blind data collector (i.e., naïve to the intent of the study) and I applied the height overlay image to each video to determine eligible focal children. Shorter video clips were created for each selected visitor group, and an Excel log describing the group, including gender and identifiable clothing, were compiled throughout the selection process. Gender was determined visually, using sex as an approximation of gender. Whenever a data collector was uncertain about an eligible child's gender, at least two additional researchers discussed the child in question. If any uncertainty remained, that child was dropped from the data set (this occurred for fewer than five visitors).

Tallies of target child gender and caregiver interaction were kept in order to obtain equal sample sizes. Given that adult/child visitor groups were the least common at the exhibit, but were the most critical group to reaching the required sample size for the study, any adult/child group was chosen over a single user during the Duration and Social Interaction selection. A running tally was kept of the number of boys and girls, single or with an adult, within each exhibit condition selected for the research; when the tally reflected an imbalance of greater than five subjects from either gender, the data collector over-sampled until the imbalance returned to five or fewer across gender. Matched days of video data were continually used for data collection until the target sample size was reached for Social Interaction data (the smallest subset).

All video data were coded using StudioCode, a software system that links a data timeline to each video clip. The software allows one to create and name code buttons, which are clicked on and off to insert code instances on the timeline. A coder can easily move forward and backward through the data at normal speed or in slow motion, which increases accuracy of start and end times for each code. The software has a function for inserting notes or labels linked to a code, which can later be referenced during coding scheme development, when discussing disagreements among coders, and when analyzing the data. All data were then exported from the StudioCode timelines into an Excel spreadsheet. Excel files were cleaned and then imported into SPSS for statistical analyses.

## Overview of Observational Codes

**Engagement.** Two common and non-obtrusive measures of visitor Engagement at an exhibit are attraction and duration (Falk, 1983). These measures were used to assess how engaging the exhibit is for girls and boys:

***Attraction.*** Attraction is operationalized in this research as the percentage of stops at the exhibit by those who enter the Sound Abatement area. I looked specifically at the percentage of girls and the percentage of boys who enter the sound-abatement area. Attraction data were collected from hour-long segments of video and entered into an Excel file. The data collector was blind to the purposes of the study. Using the height overlay image described above, the data collector tallied the number of boys and girls between ages 7 and 13 who entered the Sound Abatement area, and the number who stopped at the target exhibit (touched any exhibit component). Return visitors were not double counted; however, if the visitor had not used the exhibit during their first visit, and did use the exhibit on a return visit (within that hour), their exhibit use was recorded. See Appendix D for the Attraction coding guidelines.

***Duration.*** Duration is operationalized in this study as the amount of time spent at the exhibit, either standing directly at the exhibit and facing it after having touched at least one component, or interacting with the components of the exhibit. Time spent was recorded in seconds. If the visitor returned during the next three minutes of video data, the Duration of their second visit was also incorporated in their time spent (and on until they did not return within three minutes of their last use of the exhibit). However, if a visitor returned after three minutes had passed, the Duration of their second visit was not incorporated in their time spent (and they were not eligible for re-selection). This is because it becomes increasingly difficult to remain systematic (e.g., their first visit may be near the end of the taping session and the data may not be available), and scanning through the data adds too much additional time to the visitor group selection and data coding process. Duration video clips for Social Interaction visitors were coded by myself and two other blind coders (i.e., unaware the intent of the study). A single data collector, also naïve (blind) to the purposes of the study, coded Duration video clips for solo visitors.

The average time spent at exhibits that similarly (cf. Geometry in Motion) allow for visitor exploration of a phenomenon (as opposed to exhibits that show a single counter-intuitive result) is 3.3 minutes (Humphrey & Gutwill, 2005). The average time spent at the original version of *Geometry in Motion* was approximately 1.3 minutes.

**Social Interactions.** A coding scheme was developed to explore Social Interactions between target children and caregivers in the audio/video data. The coding scheme development and training process are described in more detail following the overview of the Social Interaction codes. The following Social Interactions were of interest to the current study (see Appendix D for the full coding scheme, including examples for each code, and Appendix E for coded excerpts from three visitor conversations):

***Kid Statements.*** *Kid Statements* entailed any utterance produced by the focal child that had to do with the exhibit, and was not clearly a question. *Kid Statements* also included content-less utterances such as grunts, sound effects and singing - all of which may be interpreted as indications of a child's Engagement or interest during the activity. However, these content-less utterances are tagged, allowing for easy removal from the analyses (i.e., this study explored their occurrence as well as typical child commentary clearly about the exhibit). Any remark or noise made by the focal child that had to do with the exhibit is a *Kid Statement*. If a remark was clearly a question, or clearly off-topic (e.g., about the museum in general or home events) it was not coded as *Kid Statement*.

***Kid Questions.*** Any question asked by the focal child that had to do with the exhibit was coded as a *Kid Question*. Off-topic Questions, such as those relating to the museum in general or to home events, were not coded. For example, "what are you supposed to do here?" or "how does it turn?" would be coded as a *Kid Question*, while "can we go to the museum store?" would not be coded.

**Adult Informative Talk.** *Informative Talk* from adults to the focal child included any information about the exhibit, exhibit use, and exhibit phenomenon directed to the child. This set of codes did not include questions (e.g., "where does this piece go?") or any acknowledging, encouraging or discouraging remarks, or statements of opinion about the exhibit (e.g., "we can get this" or "you'll like this one"). Adult Informative Talk contained three sub-codes:

**1.** ***Describe/Name***: Included verbal references to the graphics or parts, either by describing or naming, without any elaboration. This code also included statements or descriptions about what could be observed at the exhibit (referencing visual, auditory, or tactile information), with no depth or substance about "how to," "why" or analogous connections.

**2.** ***Directions/Procedure Explanations***: Included utterances providing information about how to/how not to use the exhibit or what to do/what not to do, including quick directives, goal-setting (unless they used a metaphor or analogy, which would be coded as Meaningful Explanation described below), and acknowledgement of what does/doesn't work and the building process. In sum, this code included talk about one's own or others: doing (e.g., what to do, or what one did), using (e.g., how to), trials (i.e., attempts and their outcomes), or process (e.g., steps followed or discovered). While this type of talk often involved direction commenting and procedure commenting about use at the exhibit, it did not include information about why to use the exhibit a particular way.

**3.** ***Meaningful Explanation***: Included statements providing information about cause and effect; relationships to other exhibits or other/broader phenomena; relationships to more general principles; or use of deep or superficial metaphors or analogies (such connections to pre-existing ideas aid children in making meaning of an interaction). This type of talk often provided information about why something did or did not work.

**Adult Questions.** Questions about the exhibit, posed to the focal child by the adult, were also coded. This study tracked two types of adult Questions, closed and open, to explore whether the breakdown of adult Questions seemed meaningful. *Closed-ended Questions* were questions that allowed for yes, no, or maybe responses. *Open-ended Questions* were questions that allowed for more elaborate responses than yes, no, or maybe.

Confidence in video data research relies heavily on the validity and reliability of the coding scheme. The validity of the codes used in this research is evidenced, in part, by the face validity imparted by the co-development process and the adaptation from previous museum studies schemes. Consistency across coders (inter-rater reliability) when applying the coding scheme to the video data provides support for the strength of the scheme.

**Coding Scheme Development**

In a recent article, Salinger, Planka, and Prechelt (2008) describe their use of grounded theory as a basis for coding scheme development, along with the pitfalls they experienced. The pitfalls, which resonated based on work I had done for other Geometry Playground projects, included:

> 1. The lack of predefined focus led to far too many codes, which became overwhelming and led to mistakes in judging the importance of some codes to the research question.
> 2. The lack of predefined grain size (i.e., unit of analysis; a smaller grain size equates to increased observational resolution; Nesbit & Hadwin, 2006; Roscoe, 2008) led to incomparable codes.
> 3. The lack of predefined levels of acceptable subjectivity led to too much variety in what observations could and could not be coded.
> 4. The lack of predefined concept groupings let to overlapping categories with

unclear boundaries.

The authors provide a list of suggested changes to the coding scheme development process, which I followed during coding scheme adaptation and further development, including:

**1. Identify the focus from the start.** A great many interesting behaviors occur; there is a need to identify what behaviors are most important. I addressed this suggestion by identifying the focus of my study early on, by stating my specific hypotheses and by finding pre-existing coding schemes to help narrow my observations of the video data. This suggestion also helped me to reign in my focus when I began to get excited about more and different behaviors while viewing the pilot videos.

**2. Come up with concept name syntax rules.** The name of the code shapes the data. In the course of this research, the code names have changed numerable times. For example, prior to collecting data, I dubbed adult *"Directions/Procedure"* as "how to" and later as "orientation." I continued to develop the code names after the data were collected (see pair coding below). The process of identifying boundaries between codes involved identifying inclusion and exclusion criteria; the code names began this process, but pair coding is where code names became finalized and the code boundaries were truly established.

**3. Pair coding.** The most important advice given by the authors was to use a subset of the data for pair coding, or coding done by two people working together at one computer. This process requires a consensus of two people for all important decisions: which phenomena arise in the data and into which existing code it fits; when a new code needs to be created; defining the onset of a code and determining when the code ends; naming or renaming codes; and creating and refining definitions of codes. In order to develop the coding scheme, I set aside 5% of the data for pair coding scheme development, which I later decided to maintain as pilot rather than final data. I worked as one of the pair coders. The other pair coder had previously worked in child education and in the Visitor Research & Evaluation department as an on-call data collector for about a year, and is a mother (which gave her an important perspective divergent from my own).

Several iterative steps were followed when working with the pair coder to develop the scheme. We began by thoroughly discussing the aims of the research project and the first draft of the coding scheme. The pair coder was given leeway and explicitly asked to use her knowledge base and understanding of the study to contribute to the development of the coding scheme. We then sat together and watched several videos, while taking notes. After watching the videos, we discussed our notes, referenced the study goals, made changes to the scheme, and watched a few more videos, again taking notes. After discussing our additional notes, and making corresponding changes to the scheme, we moved into the double-coding stage, independently coding the same small set of videos.

During the double-coding stage (two people coding the same video), we used a function in the StudioCode software which allows for two coded timelines to be attached to the same video. We could then compare our independent codes to identify areas of weakness in the scheme, and further define the codes. The timeline visualizations of each person's codes also allowed us to more easily operationalize the onset and termination criteria for each code. In order to identify the areas of strength and weakness in the scheme, we employed three code labels:

- Note- used to transcribe the quote and note whatever questions arose for that particular instance;
- Prototypical- used to identify statements that each pair coder considered prototypical examples of the code; and
- Borderline- used to identify instances for which inclusion or exclusion is unclear under the current scheme.

These three code labels helped drive conversations with the pair coder about possible new codes, code definitions, code inclusion and exclusion criteria, agreement on prototypical code examples, and issues regarding the scheme framing and assumptions. We repeated this process for three months (across 17 videos) until we reached consensus on the coding scheme and had created a good set of examples for each code. The entire pair coding process culminated in a coding scheme manual used to train and inform the blind coders.

**Coding Scheme Training**

Two coders in addition to myself and the pair coder were trained on the coding scheme and the StudioCode system; both were blind to the purpose of the research. Both coders have a background in social science research (one in the business setting, the other in developmental psychology), had worked in museums (including the Exploratorium) in an evaluation or research capacity, and had been involved with coding scheme development and data coding (one in regard to written responses, the other in regard to video data). Training was conducted using pilot video data, including several of the pilot videos used during coding scheme development. Training for each coder involved a month-long, seven step process, described in Table 4 below.

Table 4

*Coder training steps*

| | |
|---|---|
| Step 1 | Careful reading and discussion of the coding scheme manual; |
| Step 2 | Viewing two visitor groups' interactions at one of the focal exhibits; |
| Step 3 | Viewing of the same visitor groups from step two, this time with master codes (codes created and reviewed by the pair coders following the completion of the pair coding process) and discussion; |
| Step 4 | Real-time co-coding and discussion of a visitor group combined with initial StudioCode training; |
| Step 5 | Independent coding of a visitor group, followed by review with the master coded video (codes created by the pair coders following completion of the coding scheme development process) and code discussions; |
| Step 6 | Completing the same code, label, compare and discuss process as the pair coders (detailed above); however, codes were not changed based on these discussions. Clarifications were added to the manual during training, as well as addendums during coding; and |
| Step 7 | Double-coding an additional subset of 5% of the data until at least 90% agreement was reached for all codes. These reliability training timelines were discussed between the two coders and adjusted according to those discussions so they could be referenced throughout the coding process; however, these timelines were not maintained as data for the study. |

Training for the Attraction data followed a different trajectory. The data collector, blind to the purposes of the study, had completed the visitor group Selection process, using the same scheme that was used to record Attraction data. At the end of the Selection process, we then trained on the Attraction process. Training focused on the differing data entry compared to Selection. To ensure the most accurate data possible, we began by reaching 90% agreement on a subset of data. We each completed Attraction coding for the same two 15-minute video segments; for both segments we reached over 90% agreement. The blind data collector completed all Attraction coding and data entry.

**Coding Assignments**

After trained coders reached agreement with myself and/or each other on Duration and Social Interactions, they began coding their own videos. Videos for children with their caregivers were randomly assigned to each coder by a research assistant who was not coding the data. From the 342 assigned files, 19 were dropped due to poor audio or an inability to distinguish speakers or targets within the visitor group (6%). Thus, 323 children and their caregivers were included in the final data set. An additional 260 children who visited the exhibit alone were included in the data set to assess Engagement (Duration) for children regardless of caregiver status; because Duration coding was simple and agreement very high (99.99%), these visitor groups were not randomly assigned to the two coders (including myself) who coded this data.

A portion of the Social Interaction visitor groups (children with their caregivers) videos were further randomly assigned and co-coded to determine inter-coder reliability (described below). Reliability assessment provides evidence that coder training was well implemented and helps instill confidence in the application of the scheme across coders.

**Reliability**

Reliability coding was conducted for 19% of the Duration and Social Interaction data. Reliability videos were randomly selected from the entire dataset by assigning random numbers to each visitor group and selecting the 20% with the highest randomly assigned numbers for reliability coding (two of the co-coded videos were dropped for audio purposes, reducing the number from 20% to 19%). A research assistant who was not coding the data distributed the reliability videos throughout coder files; coders did not know which videos were reliability videos. Video data was assigned in fourteen batches; each batch contained an unknown number of reliability videos. In order to ensure stringent coding practices and reduce the amount of coder drift regarding the definition of each code throughout the entirety of the coding process, coders completed batches of videos at the same time and then discussed each subset of reliability videos before continuing to the next batch of videos. Coders were allowed to change their codes based on discussions regarding code disagreements (i.e., they created consensus codes); however, their initial codes were retained for calculating reliability. Coders were asked to discuss any confusing instances in the remaining (non-reliability) videos to ensure that all data were of the same quality as the reliability data; they typically discussed 3% of the coded instances.

In assessing reliability, transcription and coding disagreements needed to be disentangled, a process unique to coding software such as StudioCode. The StudioCode software permits researchers to watch and listen to the videotaped interactions while concurrently entering codes directly onto a timeline that runs the length of the video. The codes on each video's timeline comprise the data that can then be downloaded directly into Excel. This in-situ coding enables researchers to code video without the additional transcription step typically applied to verbal interactions, because coders can rewind and slow the speed of the verbal communication while watching physical interactions.

However, skipping verbal transcription allows disagreements between coders to arise not only from codes, but also from transcription-type decisions about who is speaking, to whom they are speaking, or what is being said. For this study, coders tracked coding scheme disagreements and transcription-type disagreements separately. The separation of these two types of disagreements allowed for a more accurate measure of coding scheme reliability, rather than one inflated by transcription reliability, which is not typically tracked or reported.

Inter-rater reliability was assessed using intraclass correlations (ICC), to assess the effective reliability of the measurement model, when possible, and percentage agreement to assess inter-rater agreement when ICCs were not possible (Rosenthal & Rosnow, 2008). The Duration data were collected such that coder agreement for the amount of time visitors were not using the exhibit could not be assessed. The non-coded data is an essential component to the ICC equation, thus, percentage agreement is reported. Coders were considered in agreement if they agreed second-by-second throughout the visitor clip; disagreements were not counted if they were within two-seconds (typical in behavioral research; see Bakeman & Gottman, 1997). Percentage agreement, the number of second-by-second agreements divided by the total number of coded seconds (Miles & Humberman, 1994) for the Duration code was 99.99% for the children who visited with caregivers; the extremely high Duration agreement for children with caregivers provided ample evidence that reliability coding was not necessary for the Duration data for children without caregivers.

All coding was conducted using StudioCode software, which allows coders to attach codes to video in real-time without pre-identifying utterances or transcribing those utterances. Therefore, coder disagreements include disagreements in four realms:

1. Code assignment (the measure used to estimate inter-rater reliability),
2. Speaker (especially when multiple children are in the group),
3. Target of the utterance (i.e., who the focal person directing their talk to), and
4. Wording (the statement content, which can be especially difficult to understand in the busy museum setting).

In an effort to disentangle the transcription-based disagreement (speaker, target, and wording) from the code assignment disagreement, coders tracked these disagreements separately. Percentage agreement for transcription (that is, speaker, target, and wording were all in agreement) was 76%. It is difficult to judge whether this is a reasonable level of transcription agreement because researchers do not typically assess or report transcription agreement. Once transcription errors were removed, ICCs for each of the Social Interaction variables were conducted.

Adult Informative Talk was originally coded as a three-level variable and later collapsed into two levels because the first level occurred for fewer than 15% of the cases and was not theoretically driven. The inter-rater reliability for Adult Informative Talk was assessed at the level at which the data were collected (three levels: Describe/Name, Directions/Procedure, and Meaningful Explanations); the ICC was .78, at the lower edge of acceptable reliability (which Shrout & Fleiss, 1979, suggest should be between .75 and .80). The ICC for Adult Questions measured coder consistency between Open-ended

Questions and Closed-ended Questions, and was well above the acceptable level at .96, The ICC for Kids' Talk measured coder consistency between Kid Statements and Kid Questions, and was .94, also well above the acceptable minimum level for reliability.

# Chapter 4: Results

## Analyses

The principal purpose of this research was to determine whether adding female-friendly design features to an exhibit enhances girls' Engagement and Social Interactions at the exhibit. Ancillary to the main research goal was the desire to ensure that the female-friendly design features have no undesirable effects on boys' Engagement and Social Interactions at the exhibit. According to Rosnow and Rosenthal (1996), researchers should conduct analyses (contrasts or interactions) that embody the theory, hypothesis, or question of interest. These authors, with support from Abelson (1996), suggest that researchers must conduct analyses that answer the question at hand and support the point they are (thoughtfully) attempting to make. Because the primary focus of this research pertains to the effects of female-friendly design on girls, analyses focus on investigating the simple effects of exhibit type for girls. Subsidiary analyses investigating the simple effects of exhibit type for boys follow. Finally, counter to Rosnow and Rosenthal's (1996), and Rosenthal and Rosnow's (2008) suggestions that analyses of interactions should not be conducted unless they are the focus of the research, because they increase family-wise error (i.e., likelihood of making a type one error), I follow-up simple effects with interaction analyses to determine whether any additional information can be gained when considering these two simple effects relative to one another. While the order of analyses undertaken for the current research is atypical, it pointedly answers the research questions and hypotheses in a manner supported by the goals of the research as well as psychological and mathematical statisticians (Abelson, 1996; Rosenthal & Rosnow, 2008; Rosnow & Rosenthal, 1996).

Many of the analyses below assume normality; however, the data from this study are highly negatively skewed (i.e., rather than normally distributed, the majority of cases are in the lower end of the scale, with only a few cases reaching the higher ends of the scale). In these cases, sensitivity analyses were conducted to determine whether the mean difference results substantively changed when:

- the top 5% and bottom 5% tails were trimmed,
- the data were logarithmically transformed (after adding a constant),
- the Social Interactions were normalized by the time spent , or
- non-parametric analyses were used (that is, tests of rank rather than mean).

For each of the variables, most of the sensitivity analyses were consistent with the mean differences. Below I report the mean differences per hypothesis, and footnote any results that were contrary to the reported results. All analyses were two-tailed.

**Research question 1.** Does girls' Engagement with the exhibit (Attraction and Duration) depend on whether female-friendly design features are incorporated?

> ***Hypothesis 1a.*** It was predicted that more girls who entered the Sound Abatement area would stop at *Geometry In Motion* when the Female-Friendly design Features were added to the exhibit. A chi square test of independence was conducted to determine whether girls are more likely to stop at the exhibit once the design features are incorporated. Of the girls who entered the sound abatement area, those who entered when the female-friendly version of the

exhibit was present were more likely to stop at the exhibit (61%) than those who entered when the Non-Featured version of the exhibit was present (45%), $X^2(1, N = 160) = 4.24, p = .04$). See Table 5. Girls were more attracted to the Female-Friendly Featured version of the exhibit, indicating that there is a strong benefit to including the female-friendly design features.

***Hypothesis 1b***. It was predicted that boys who entered the Sound Abatement area would stop at *Geometry In Motion* equally as often when the Female-Friendly design Features were added to the exhibit. A chi square test of independence was conducted to determine whether boys are equally likely to stop at the exhibit regardless of the design features. Of the boys who entered the sound abatement area, those who entered when the female-friendly version of the exhibit was present were more likely to stop at the exhibit (62%) than those who entered when the Non-Featured version of the exhibit was present (47%), $X^2(1, N = 185) = 4.34, p = .04$. See Table 6. Boys were more attracted to the Female-Friendly Featured exhibit, indicating that the female-friendly design features do not negatively impact boys' attraction to the exhibit.

***Hypothesis 1 follow-up.*** A direct logistic regression was conducted to determine whether the higher Attraction rate (more stops at the female-friendly version compared to the Non-Featured version) is greater for girls than for boys. Contrary to expectations, girls' increases in Attraction were not significantly different from boys' increases in Attraction once the female-friendly features were incorporated (compared to the Non-Featured exhibit): Gender*Exhibit interaction slope = .04, Wald $Z < .01$, $p = .93$. See Tables 5 and

6. In sum, both boys and girls were significantly more attracted to the exhibit once the Female-Friendly design Features were added.

Table 5

*Counts of girls who stop at the exhibit as a function of female-friendly design features*

| Girls | Stop | Do not stop |
|---|---|---|
| Female-Friendly Featured version | 49 | 31 |
| Non-Featured version | 36 | 44 |

Table 6

*Counts of boys who stop at the exhibit as a function of female-friendly design features*

| Boys | Stop | Do not stop |
|---|---|---|
| Female-Friendly Featured version | 60 | 37 |
| Non-Featured version | 41 | 47 |

***Hypothesis 2a***. It was predicted that girls would spend more time at the Female-Friendly Featured exhibit than at the Non-Featured exhibit. An initial t test across both versions of the exhibit revealed that the Duration data in minutes for girls without caregivers ($M = .79$, $SD = 1.58$, $n = 122$) and girls with caregivers ($M = 2.57$, $SD = 2.78$, $n = 157$) were significantly different and could not be combined, $t(256) = -6.74$, $p < .01$. Independent samples t tests were conducted to determine whether girls without caregivers and girls

who visit with caregivers spend significantly more time at the exhibit with female-friendly design features incorporated. Counter to expectations, girls without caregivers did not spend significantly less time at the Non-Featured exhibit (M = .88 minutes, SD = 2.06, n = 65) than those at the Female-Friendly Featured exhibit (M = .68 minutes, SD = .73, n = 57), t(120) = .70, p = .49, *d* = .13. However, in accordance with expectations, when girls visited the exhibit with an adult, they spent significantly more time at the Female-Friendly Featured exhibit (M = 2.97 minutes, SD = 3.24, n = 83) than at the Non-Featured exhibit (M = 2.12, SD = 2.08, n = 74), t(142) = -1.99, p = .05, *d* = .32[4]. See the girls' Durations in Table 7. When girls visit the exhibit with caregivers, they spend more time at the Female-Friendly Featured version, indicating that there is a strong benefit to including such design features.

***Hypothesis 2b***. It was predicted that boys would spend similar amounts of time at the Female Friendly Featured exhibit than at the Non-Featured exhibit. An initial t test across both versions of the exhibit revealed that the Duration data in minutes for boys without caregivers (M = .83, SD = 1.37, n = 138) and boys with caregivers (M = 3.94, SD = 5.90, n = 166) were significantly different and could not be combined, t(186) = -6.59, p < .01. Independent samples t tests were conducted to determine whether boys who visit without caregivers and boys who visit with caregivers did spend similar amounts of

[4] While all of the sensitivity results were in the same direction, two of the sensitivity analyses did not yield significant results: Mann Whitney U: Z = -1.47, p = .14, and the logarithmic transformation: $t_{log}(154) = -1.71$, p = .09.

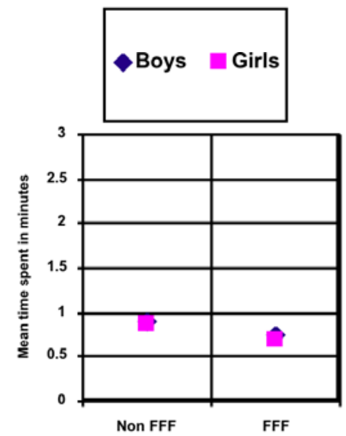
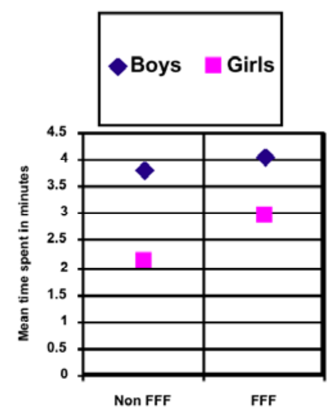
time at the exhibit regardless of the design features. Concordant with expectations, boys who visited the exhibit without caregivers did not spend significantly less time at the Non-Featured ($M = .91$, $SD = 1.53$, $n = 69$) exhibit compared to the Female-Friendly Featured ($M = .75$, $SD = 1.2$, $n = 69$) exhibit, $t(136) = .67$, $p = .50$, $d = .11$. Similarly, boys who visited with caregivers did not spend significantly less time at the Non-Featured ($M = 3.83$, $SD = 6.31$, $n = 78$) exhibit compared to the Female-Friendly Featured ($M = 4.05$, $SD = 5.54$, $n = 88$) exhibit, $t(164) = -.24$, $p = .81$, $d = .04$. See the boys' Durations in Table 7. While no significant ill effects on time spent for boys were detected between the Non-Featured and the Female-Friendly Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power) or the particular sample.

***Hypothesis 2 follow-up.*** An initial t test across both versions of the exhibit indicated that the Duration data for children without caregivers ($M = .81$, $SD = 1.47$) were significantly different from the Duration data for children who visited with caregivers ($M = 3.28$, $SD = 4.70$) and should not be combined, $t(398) = -8.90$, $p < .01$. Two-by-two ANOVAs were conducted to determine whether a significant interaction occurred between gender and exhibit version in the Duration of time spent at the exhibit. For the children without caregivers, the difference in time spent between exhibit versions did not significantly depend on gender, $F(1, 260) = .014$, $p = .91$, partial eta-squared < .01. Likewise, for children with caregivers, the difference in time spent between exhibit versions did not depend on gender, $F(1, 323) = .34$, $p = .54$, partial eta-squared < .01. See Table 7. It appears that the increase in time spent from the Non-Featured to the Female-Friendly Featured exhibit, while significant for girls, was not significantly greater for girls than the increase for boys. However, this result is inconclusive and could be due to insufficient power (e.g., the sample size or the variance heterogeneity) or the particular sample.

In an effort to further understand the Duration results, a three-way 2X2X2 ANOVA was conducted to explore whether the Duration results depended on an interaction between caregivers' presence, child's gender and exhibit version. The positive effects of the exhibit design features on Duration results were not significantly greater for girls than boys when visiting alone, and that discrepancy was not significantly stronger when children visited with a caregiver, $F(1, 582) = .32$, $p = .57$, partial eta-squared < .01. See Table 7.

Table 7

*Means and plots Durations as a function of gender and exhibit version for children without caregivers and those with caregivers*

<table>
<tr><th>Mean Durations in minutes</th><th>Non-Featured (SD)</th><th>Female-Friendly Featured (SD)</th><th>Plots of mean Durations in minutes as a function of gender and exhibit version.</th></tr>
<tr><td>Girls without caregivers</td><td>.88<br>(2.06)<br>n = 65</td><td>.68<br>(.73)<br>n =57</td><td rowspan="2">
</td></tr>
<tr><td>Boys without caregivers</td><td>.91<br>(1.53)<br>n = 69</td><td>.75<br>(1.20)<br>n = 69</td></tr>
<tr><td>Girls with caregivers</td><td>2.12<br>(2.08)<br>n = 74</td><td>2.97*<br>(3.24)<br>n = 83</td><td rowspan="2">
</td></tr>
<tr><td>Boys with caregivers</td><td>3.83<br>(6.31)<br>n = 78</td><td>4.05<br>(5.54)<br>n = 88</td></tr>
</table>

*p < .05 **p < .01 ***p < .001

The pattern of the Engagement results is interesting. Girls and boys were more likely to visit the exhibit once the female-friendly design features were added, suggesting that the changes to the exhibit made it more attractive overall. Girls with caregivers also spent significantly more time at the exhibit once the female-friendly design features were added, suggesting that the female-friendly design features did make the exhibit more engaging for girls. However, for the girls without caregivers, as well as for boys regardless of caregiver accompaniment, no significant beneficial or detrimental effects were detected, leading to inconclusive results for these groups. Furthermore, the interaction analysis that explored whether the significant gains in time spent for girls was greater than the gains for boys was also inconclusive. The remainder of the analyses explored whether the verbal Social Interactions between children and their caregivers differed between the Non-Featured and Female-Friendly Featured exhibits.

**Research question 2.** Does the quality of girls' Social Interactions with caregivers at the exhibit depend on whether female-friendly design features are incorporated?

***Hypothesis 3a.*** It was predicted that girls would utter more Statements at the Female-Friendly Featured exhibit than at the Non-Featured exhibit. An independent samples t test was conducted to determine whether the number of girls' Statements was higher at the Female-Friendly Featured version of the exhibit, compared to the Non-Featured version. Contrary to expectations, results indicate that the mean number of girls' Statements did not differ significantly between the Non-Featured ($M = 7.27$, $SD = 9.29$, $n = 74$) and the Female-Friendly Featured ($M = 10.20$, $SD = 12.86$, $n = 83$) versions of the exhibit, t(149) = -1.65, p = .10, *d* = .26[5]. See the girls' number of Statements in Table 8. Girls tended to utter a greater number of Statements at the Female-Friendly Featured exhibit than at the Non-Featured exhibit, but this difference was not significant, suggesting that further research is needed.

***Hypothesis 3b.*** It was predicted that boys would utter a similar number of Statements at the Female-Friendly Featured exhibit and the Non-Featured exhibit. An independent samples t test was conducted to determine whether boys utter a similar number of Statements at the exhibit regardless of the design features. Concordant with expectations, the mean number of boys' Statements did not differ significantly between the Non-Featured (M = 11.26, SD = 15.85, n = 78) and the Female-Friendly Featured (M = 11.56, SD = 21.83, n = 88) versions of the exhibits, t(164) = -.10, p = .92, *d* = .02. See the boys' number of Statements in Table 8. While no significant ill effects on boys' statements were detected between the Non-Featured and the Female-Friendly Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power) or the particular sample.

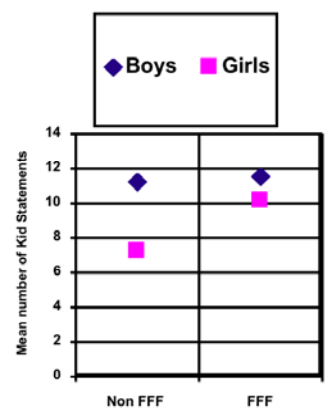
***Hypothesis 3 follow-up***. A 2X2 ANOVA was conducted to determine whether an interaction occurs between gender and the exhibit version in the number of Statements children uttered. Counter to expectations, adding female-friendly features did not lead to significantly higher increases in number of Statements

[5] Note that these results were sensitive to the distribution of the data, and while not significant here or for the logarithmically transformed, normalized, or non-parametric analyses, these results were significant when the tails were trimmed, t(121). = 12.19, p = .03. Given the sensitive nature, these results are reported as non-significant using analyses that take all participating visitor groups (i.e., all data points) into consideration.

from the Non-Featured version to the Female-Friendly Featured version for girls than for boys, F(1, 323) = .55, p = .46, partial eta-squared < .01. See Table 8. The increase in number of Kids' Statements uttered from the Non-Featured to the Female-Friendly Featured exhibits, while somewhat greater for girls than boys, was not significantly greater and warrants further research.

Table 8

*Means and plot of Kids' Statements as a function of gender and exhibit version*

| Mean number of Kids' Statements | Non-Featured (SD) | Female-Friendly Featured (SD) | Plot of mean number of Kids' Statements as a function of gender and exhibit version. |
|---|---|---|---|
| Girls | 7.27<br>(9.29)<br>n = 74 | 10.20<br>(12.86)<br>n = 83 |  |
| Boys | 11.26<br>(15.85)<br>n = 78 | 11.56<br>(21.83)<br>n = 88 | |



*p < .05 **p < .01 ***p < .001

***Hypothesis 4a.*** It was predicted that girls would ask more Questions at the Female-Friendly Featured exhibit than at the Non-Featured exhibit. An independent samples t test was conducted to determine whether girls ask more Questions at the exhibit once the design features are incorporated. Counter to expectations, results indicate that the mean number of girls' Questions at the

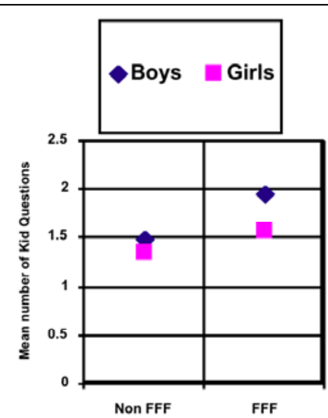
Non-Featured exhibit (M = 1.35, SD = 2.34, n= 74) did not differ significantly from the mean number of questions at the Female-Friendly Featured exhibit (M = 1.57, SD = 2.38, n = 83), t(155) = -.57, p = .57, *d* = .09. See the girls' number of Questions asked in Table 9. While girls tended to ask similar numbers of Questions at the Female-Friendly Featured and Non-Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power) or the particular sample.

***Hypothesis 4b***. It was predicted that boys would ask a similar number of Questions at the Female-Friendly Featured and the Non-Featured exhibits. An independent samples t test was conducted to determine whether boys ask a similar number of Questions at the exhibit regardless of the design features. Concordant with expectations, the mean number of boys' Questions did not differ significantly between Non-Featured (M = 1.49, SD = 2.26, n = 78) and the Female-Friendly Featured (M = 1.95, SD = 3.65, n = 88) versions of the exhibits, t(164) = -.98, p = .33, *d* = .16. See the boys' number of Questions asked in Table 9. While no significant ill effects on boys' number of Questions asked between the Female-Friendly Featured and Non-Featured exhibits were detected, these results are inconclusive and could be due to the sample size (i.e., power) or the particular sample.

***Hypothesis 4 follow-up.*** A 2X2 ANOVA was conducted to determine whether an interaction occurred between gender and the exhibit version in the number of Questions asked by children. Contrary to expectations, female-friendly design features did not lead to significantly higher increases in number of Questions asked for girls than for boys, $F(1, 323) = .17$, $p = .68$, partial eta-squared < .01. See Table 9. The increase in number of Kids' Questions asked at the Non-Featured to the Female-Friendly Featured exhibit was minimal and did not differ significantly between girls and boys; however, these results are inconclusive a may be due to the sample size (i.e., power) or the particular sample.

Table 9

*Means and plot Kids' Questions asked as a function of gender and exhibit version*

| Mean number of Kids' Questions | Non-Featured (SD) | Female-Friendly Featured (SD) | Plot of mean number of Kids' Questions asked as a function of gender and exhibit version. |
|---|---|---|---|
| Girls | 1.35 (2.34) n = 74 | 1.57 (2.38) n = 83 |  |
| Boys | 1.49 (2.26) n = 78 | 1.95 (3.65) n = 88 | |

*p < .05 **p < .01 ***p < .001

The results suggest that the quality of Kids' Social Interactions with caregivers at the exhibit may not depend on whether female-friendly design features are incorporated. Specifically, adding the female-friendly features to the exhibit did not significantly affect the number of Statements children uttered, regardless of gender. This study was unable to detect any ill (or positive) effects on the number of Kid Statements for boys or girls, leading to inconclusive results. However, the pattern of girls' Statements does show a small increase in number (as indicated by the small effect size, $d = .26$), suggesting that future research is warranted. While the difference in number of Statements by girls compared to those by boys became smaller descriptively, the interaction was not significant, leading to inconclusive results. I further explored whether the definition of Kid Statements had an impact on results by removing the content-free Kid Statements (such as sound-effects, grunts, and humming) from the analyses; the results remained unchanged. Furthermore, results for Kid Questions were also inconclusive; the female-friendly design features did not significantly affect the number of Kid Questions, regardless of gender. See Appendix E for coded excerpts from three visitor conversations.

**Research question 3.** Does the quality of caregivers' Social Interactions with girls at the exhibit depend on whether female-friendly design features are incorporated?

***Hypothesis 5a.*** It was predicted that caregivers would provide girls with more *Meaningful Explanations* at the Female-Friendly Featured exhibit than at the Non-Featured exhibit. An independent samples t test was conducted to determine whether caregivers provide girls significantly more *Meaningful Explanations* once the design features are incorporated. Contrary to expectations, the mean number of caregivers' *Meaningful Explanations* did not significantly differ for girls at the Non-Featured (M = .86, SD = 1.58, n =

74) and the Female-Friendly Featured (M = .95, SD = 1.55, n = 83) versions of the exhibit, t(155) = -.35, p = .73, *d* = .06. See the *Meaningful Explanations* provided girls plotted in Table 10. While caregivers tended to provide girls a similar number of *Meaningful Explanations* at the Non-Featured and Female-Friendly Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power), the particular sample, or the operational definition for Meaningful Explanations.

***Hypothesis 5b.*** It was predicted that caregivers would provide girls with a similar number of *Direction/Procedure Explanations* at the Female-Friendly Featured exhibit and the Non-Featured exhibit. An independent samples t test was conducted to determine whether caregivers provide girls a similar number of *Direction/Procedure Explanations* at the exhibit regardless of the design features. Concordant with expectations, the mean number of caregivers' *Direction/Procedure Explanations* did not significantly differ for girls between the Non-Featured (M = 4.76, SD = 6.58, n = 74) and the Female-Friendly Featured (M = 5.25, SD = 7.55, n = 83) versions of the exhibit, t(155) = -.44, p = .66, *d* = .07[6]. See the *Direction/Procedure Explanations* provided girls' plotted in Table 10. While it seems that caregivers provide girls a similar number of *Direction/Procedure Explanations* at the Non-

[6] Note that these results were sensitive to the distribution of the data, and while not significant here or for the trimmed, logarithmically transformed, or non-parametric analyses, these results were significant when the data were normalized by time spent at the exhibit, t (124) = 2.13, p = .04. Given the sensitive nature, and the likelihood that visitors are more talkative when first learning to use the exhibit (thus normalized results are skewed in favor of the Non-Featured exhibit where girls spent less time), these results are reported as non-significant.

Featured and Female-Friendly Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power), the particular sample, or the operational definition for Direction/Procedure Explanations.

***Hypotheses 5a and 5b follow-up.*** A 2X2 repeated measures ANOVA was conducted to determine whether there is a significant interaction between the type of explanation and exhibit version on number of explanations given by the caregiver (within the sample of girls). Counter to expectations, changes to the exhibit did not have a significantly greater impact on the number of *Meaningful Explanations* by caregivers than on the number of *Directions/Procedure Explanations* by caregivers, $F(1, 155) = .18$, $p = .67$, partial eta-squared < .01. See the girls' data plotted in Table 10. The increase in number of caregivers' *Meaningful Explanations* provided to girls, compared to *Direction/Procedure Explanations*, did not differ significantly between the Non-Featured exhibit and the Female-Friendly Featured exhibit.

***Hypothesis 5c***. It was predicted that caregivers would provide boys a similar number of *Meaningful* Explanations at the Female-Friendly Featured exhibit and the Non-Featured exhibit. An independent samples t test was conducted to determine whether caregivers provide boys similar numbers of *Meaningful* Explanations at the exhibit regardless of the design features. Concordant with expectations, the mean number of caregivers' *Meaningful Explanations* did not significantly differ for boys at the Non-Featured ($M = 1.31$, $SD = 2.75$, $n = 78$) and the Female-Friendly Featured ($M = 1.02$, $SD = 1.76$, $n = 88$) versions of the exhibits, t(164) = .805, p = .42, *d* = .13. See the *Meaningful Explanations* provided boys' plotted in Table 10. While caregivers tended to provide boys a similar number of *Meaningful Explanations* at the Non-Featured and Female-Friendly Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power), the particular sample, or the operational definition for Meaningful Explanations.

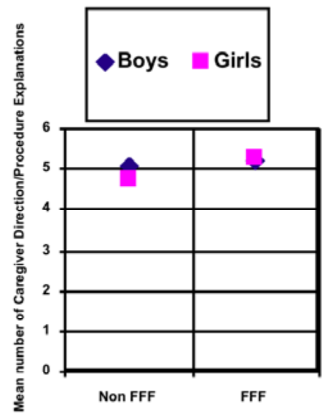
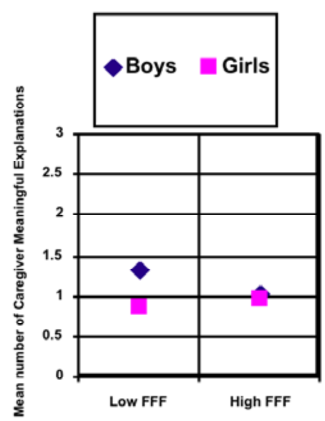
***Hypothesis 5d.*** It was predicted that caregivers would provide boys a similar number of *Direction/Procedure Explanations* at the Female-Friendly Featured exhibit and the Non-Featured exhibit. An independent samples t test was conducted to determine whether caregivers provide boys a similar number of *Direction/Procedure Explanations* at the exhibit regardless of the design features. Concordant with expectations, the mean number of caregivers' *Direction/Procedure Explanations* did not differ significantly for boys between the Non-Featured (M = 5.05, SD = 8.01, n = 78) and the Female-Friendly Featured (M = 5.22, SD = 7.16, n = 88) versions of the exhibit, t(164) = -.14, p = .89, *d* = .02. See the *Direction/Procedure Explanations* provided boys' plotted in Table 10. While it seems that caregivers provide boys a similar number of *Direction/Procedure Explanations* at the Non-Featured and Female-Friendly Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power), the particular sample, or the operational definition for Direction/Procedure Explanations.

***Hypotheses 5c and 5d follow-up.*** In accordance with the proposed analyses, a 2X2 repeated measures analysis was not conducted because hypotheses 5c and 5d were supported. That is, because there were no differences in *Meaningful Explanations* or *Direction/Procedure Explanations* between the exhibits for boys, I did not further explore whether an interaction exists between the type of explanation and exhibit version on number of explanations given by the caregiver (within the sample of boys).

***Hypothesis 5 follow-up.*** A 2X2X2 repeated measures factorial ANOVA was conducted to assess whether the number of explanations provided by caregivers depends on an interaction between type of explanation, child's gender and exhibit version. The within-subjects factor was type of explanation (Meaningful or Directions/Procedure), and the between-subjects factors were child's gender (boy or girl) and exhibit version (Non-Featured or Female-Friendly Featured). Contrary to expectations, the positive effects of the exhibit design features (condition) on number of caregiver explanations were not greater for girls than boys in the *Meaningful Explanations*, and that discrepancy was not stronger than in the *Directions/Procedure Explanations*, $F(1, 319) = .001$, $p = .98$, partial eta-squared < .01. See Table 10.

Table 10

*Means and plots of caregiver Explanations as a function of gender and exhibit version*

| Mean number of caregiver Explanations by type | Non-Featured (SD) | Female-Friendly Featured (SD) | Plots of mean number of caregiver Explanations by type as a function of gender and exhibit version. |
|---|---|---|---|
| Directions/ Procedure Explanations to Girls | 4.76 (6.58) n = 74 | 5.25 (7.55) n = 83 |   |
| Directions/ Procedure Explanations to Boys | 5.05 (8.01) n = 78 | 5.22 (7.16) n = 88 | |
| Meaningful Explanations to Girls | .86 (1.58) n = 74 | .95 (1.55) n = 83 |   |
| Meaningful Explanations to Boys | 1.31 (2.75) n = 78 | 1.02 (1.76) n = 88 | |

*p < .05 **p < .01 ***p < .001

***Hypothesis 6a.*** It was predicted that caregivers would ask girls more Questions at the Female-Friendly Featured exhibit than at the Non-Featured exhibit. An independent samples t test was conducted to determine whether caregivers ask more Questions of their girls once the design features are incorporated. Expectations were not confirmed; the mean number of Questions asked by caregivers of their girls did not differ significantly between the Non-Featured (M = 2.05, SD = 3.83, n = 74) and the Female-Friendly Featured (M = 2.18, SD = 5.24, n = 83) versions of the exhibit, t(155) = -.17, p = .86, $d$ = .03[7, 8]. See the Questions asked of girls plotted in Table 11. Caregivers tended to ask girls a similar number of Questions at the Female-Friendly Featured and the Non-Featured exhibits; however, these results are inconclusive and could be due to the sample size (i.e., power), or the particular sample.

***Hypothesis 6b.*** It was predicted that caregivers would ask boys a similar number of Questions at the Female-Friendly Featured exhibit and the Non-Featured exhibit. An independent samples t test was conducted to determine whether caregivers ask a similar number of Questions of their boys regardless of the design features. Concordant with expectations, the mean number of

---

[7] Note that these results were sensitive to the distribution of the data, and while not significant here or for the trimmed, logarithmically transformed, or non-parametric analyses, these results were significant when the data were normalized by time spent at the exhibit, t(121) = 2.00, p < .05. Given the sensitive nature, and the likelihood that visitors are more talkative when first learning to use the exhibit (thus normalized results are skewed in favor of the Non-Featured exhibit where girls spent less time), these results are reported as non-significant.

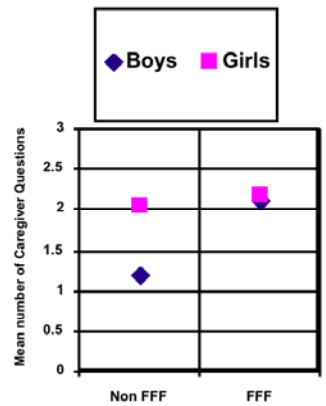
[8] Question type (open- and closed-ended) was also explored for hypotheses 6a and 6b; however, the results did not differ. That is, the mean number of open-ended and closed-ended questions (asked of girls or boys) did not differ significantly between the two versions of the exhibit.

Questions asked by caregivers of their boys was similar at the Non-Featured (M = 1.62 SD = 2.29, n = 78) and the Female-Friendly Featured (M = 2.10, SD = 4.25, n = 88) versions of the exhibit, t(164) = -.88, p = .38, *d* = .14. See the Questions asked of boys plotted in Table 11. While caregivers tended to ask boys a similar number of Questions at the Female-Friendly Featured and the Non-Featured exhibits, these results are inconclusive and could be due to the sample size (i.e., power), or the particular sample.

***Hypothesis 6 follow-up.*** A 2X2 ANOVA was conducted to determine whether an interaction occurs between child's gender and exhibit condition in the number of Questions caregivers ask. Counter to expectations, the design features did not lead to significantly higher increases in number of Questions asked of girls than in the number of Questions asked of boys, F(1, 323) = .15, p = .70, partial eta-squared < .01. See Table 11. The increase in number of caregivers' Questions asked at the Non-Featured to the Female-Friendly Featured exhibit did not differ significantly between girls and boys; however, these inconclusive results may also be an artifact of insufficient power or the particular sample in the study.

Table 11

*Means and plot of caregiver Questions as a function of child's gender and exhibit version*

| Mean number of caregiver Questions | Non-Featured (SD) | Female-Friendly Featured (SD) | Plot of mean number of caregiver Questions asked as a function of gender and exhibit version. |
|---|---|---|---|
| Girls | 2.05<br>(3.83)<br>n = 74 | 2.18<br>(5.24)<br>n = 83 |  |
| Boys | 1.62<br>(2.29)<br>n = 78 | 2.10<br>(4.25)<br>n = 88 | |



*p < .05 **p < .01 ***p < .001

These results suggest that the addition of female-friendly features did not significantly affect caregivers' verbal Social Interactions with their children, leading to inconclusive results. Specifically, adding the female-friendly features to the exhibit did not significantly affect the number or type of explanations caregivers provided for their children (Meaningful or Directions/Procedure), regardless of gender. I further explored whether the definition of Adult Informative Talk (Meaningful and Directions/Procedure) had an impact on results by removing the explanations caregivers read from the exhibit graphics; the results remained unchanged.

Furthermore, adding female-friendly design features did not significantly affect the number Questions caregivers asked of their children, regardless of gender. See Appendix E for coded excerpts from three visitor conversations.

Overall, the findings from this study suggest that adding female-friendly design features can enhance girls' experience at an exhibit. The Engagement results for this study are particularly positive. Girls were significantly more attracted to the exhibit once the design features were added. Girls also spent significantly more time at the Female-Friendly Featured exhibit than at the Non-Featured exhibit, when visiting with a caregiver.  While girls' Statements were not significantly higher at the Female-Friendly version of the exhibit, the notable effect size indicates that further research with higher power may be able to detect a significant effect. The remainder of the Social Interaction results for girls were inconclusive, and will be explored in more detail in the discussion section. While it is encouraging that this study was unable to detect any negative results for boys, such inconclusive results should not dissuade design teams who incorporate female-friendly design features in exhibits, and researchers, from continuing to check for unanticipated ill effects for boys. All of these results along with more detailed interpretations, limitations, and implications are discussed below.

## Chapter 5: Discussion

This research investigated the behaviors of science museum visitors at two versions of a geometry exhibit, one of which was specifically designed to increase engagement for girls and to encourage more equitable social interactions with their caregivers. The primary focus of this research was to determine whether the female-friendly design features enhanced girls' experiences at the exhibit. Subsidiary to the main research goal was the desire to ensure that the female-friendly design features had no undesirable effects on boys' engagement and social experiences at the exhibit. The social interaction component of this research is a conceptual replication of Crowley's Power Girl study (personal communication, October, 22, 2007). Crowley's earlier work identified a gender discrepancy in the way parents socially interact with their girls at science exhibits (2001), and the follow-up Power Girl study found that including a female mascot in the graphics can significantly reduce that discrepancy. The three overarching research questions for the current study were:

1. Does girls' Engagement with the exhibit depend on whether female-friendly design features are incorporated?
2. Does the quality of girls' Social Interactions with caregivers at the exhibit depend on whether female-friendly design features are incorporated?
3. Does the quality of caregivers' Social Interactions with girls at the exhibit depend on whether female-friendly design features are incorporated?

The results of this study are discussed in terms of support or failure to reinforce each research question. Each finding is followed by potential explanations and suggestions for future research. The discussion section concludes with careful consideration of the strengths, limitations, and implications of the research study.

**Engagement**

The pattern of the Engagement results, evidenced by exhibit use (Attraction) and time spent (Duration) results, is interesting. Girls and boys were more likely to visit the exhibit once the female-friendly design features were added, and girls' increased Attraction was not greater than boys'. Therefore, the female-friendly design features were able to enhance girls' Attraction to the exhibit and did not produce ill-effects to boys' Attraction to the exhibit. In fact, these results suggest that the changes to the exhibit made it more attractive overall. One explanation for the increased overall attractiveness is that the female-friendly changes that make the exhibit more attractive to females are similarly attractive to males; gender differences may be less prominent in the area of exhibit attraction (at least for this particular exhibit). Another possible explanation for increases in attractiveness for girls and boys is that the inclusion of multiple stations, one of the design changes to the Female-Friendly Featured exhibit, also increased the availability of an unoccupied station and simply made it possible for more visitors to use the exhibit. Multiple stations have been shown to increase engagement at exhibits for family groups (Borun & Dritsas, 1997; Borun et al., 1998; Humphrey & Gutwill, 2005). However, many visitors approached a station while others were using it and shared a station with other visitors, making it unlikely that the multiple stations fully explain the strong Attraction findings. Future researchers will want to disentangle the results for each of the female-friendly design features to determine whether each change affects girls' and boys' Attraction similarly.

While Attraction data included children who visited the exhibit alone combined with those who visited with caregivers, the Duration data for these two groups required separation because girls and boys with caregivers spent significantly more time at the exhibit than girls and boys without caregivers. This result was unanticipated; however, it is partially supported by a museum study that found daughters to become more engaged at exhibits when working with another family member (Diamond, 1994). This finding is also supported by the fact that a subset of the female-friendly design changes, such as multiple stations and relevant connections to everyday life, have been shown to increase time spent by families at exhibits (Borun & Dritsas, 1997; Borun, et al., 1998; Humphrey & Gutwill, 2005). These findings lend support to the developmental Contextualist perspective that engagement in learning activities is more likely when an adult is present (Miller, 1993).

The Duration data indicate that the female-friendly changes had more focused effects on girls, compared to exhibit Attraction. While this study was unable to detect any significant effects on time spent when girls were visiting alone, girls did spend significantly more time at the exhibit once the female-friendly changes were made when visiting with a caregiver. These results suggest that the female-friendly design features positively affect the amount of time girls spend at the exhibit when with a caregiver, and provide ample support for considering female-friendly design features when developing potentially inequitable exhibits. It is possible that the changes to the exhibit only affect dynamic interactions between girls and their caregivers with the exhibit. A follow-up analysis does not support this claim; however it is possible that the non-significant interaction between exhibit version and caregiver presence for girls was an artifact of sample size, $F(1, 278) = 3.56$, $p = .06$, partial eta-squared = .01. Future research should continue to explore whether there is an interaction based on caregiver presence, which would be partially explained by the Bronfenbrennarian developmental systems theoretical perspective that (learning) behaviors are more transactional than linear (1998). The results for girls with caregivers supports the notion that something transactional may be occurring, similar to the conclusions drawn in the Power Girl study, where adding a female character to the label led to increased scaffolding for girls by their parents (K. Crowley, personal communication, October 22, 2007;Crowley, 2001; Schneider & Cheslock, 2003). These results also support Social Contextualist theory such that deeper engagement occurred when interacting with an adult in the learning environment (Miller, 1993; Vygotsky, 1978).

This study was unable to detect any significant ill-effects of the female-friendly design features on boys' time spent; while encouraging, these results do not conclusively establish the absence of negative effects on boys. It is possible that there are negative effects that were not measured, that there were negative (or positive) effects that were too small to detect with the given sample size, or that the particular sample chosen for the study was aberrant. According to West et al. (2000), one can use such findings to support extra-statistical claims of the discriminant validity of a study; that is, the changes to the exhibit affected girls, but did not generalize to boys who were outside of the theoretical expectations. Unfortunately, the interaction analysis aimed at testing this interpretation was not significant, leading to inconclusive results. Perhaps the study did not have sufficient power to determine whether the results were stronger for girls, or perhaps the exhibit was simply a better exhibit enhancing engagement for girls as well as boys. Follow-up analyses determined that there were significant gender discrepancies when caregivers were present at the Non-Featured version of the exhibit in favor of males, $t(94) = 2.27$, $p = .03$, and no significant difference between girl's and boys' time spent with caregivers were detected at the Female-Friendly Featured version of the exhibit $t(142) = 1.56$, $p = .69$. Unfortunately, the aforementioned interaction analysis indicates non-significant, inconclusive results: the gender gap was not significantly reduced and further research is needed to determine whether the female-friendly design features can help reach gender equity at discrepant exhibits. The remainder of the analyses explored whether the verbal Social Interactions dynamic between children and their caregivers differed at the Non-Featured and Female-Friendly Featured exhibits.

**Girls' and Boys' Social Interactions with Caregivers**

Contrary to expectations, the quality of girls' Social Interactions with their caregivers did not fundamentally differ between the Non-Featured and the Female-Friendly Featured versions of the exhibit. Previous researchers found that adults' explanations directed to their girls became more meaningful once a female mascot was added to the labels (K. Crowley, personal communication, October 27, 2007; Crowley, 2001). These researchers did not explore girls' Statements and only examined girls' Questions in the 10-seconds, and 60-seconds, preceding an adult explanation (Crowley, et al., 2001). The current study aimed to learn more about girls' contributions to parent-child interactions by including Kid Statements and looking more broadly at Kids' Questions overall. The average number of girls' Statements was greater at the Female-

Friendly Featured version than at the Non-Featured version of the exhibit as indicated by the notable effect size (d = .26); however, this difference was not significant. The girls' Statement results suggest that the sample size was too small to detect a significant difference between the Non-Featured and the Female-Friendly Featured exhibit versions; future research should continue to explore this potential effect of female-friendly design features.

This study did not detect any significant ill effects for boys' Statements due to the female-friendly features; however, it is unclear whether the lack of significant ill effects is due to a lack of negative impacts or due to insufficient power or the particular sample in the study. The interaction between gender and exhibit version on Kid Statements was not statistically significant, and therefore inconclusive. However, the pattern of the data is promising for museum practitioners and warrants future research with a larger sample size: the number of girls' Statements were, on average, more similar to the average number of boys' statements at the Female-Friendly Featured version than the Non-Featured version. An additional analysis was conducted to determine whether there was a significant disparity between boys' and girls' number of statements at the Non-Featured version of the exhibit; while not significant, $t(169) = 1.90$. $p = .06$, the effect size is notable ($d = .31$); future research should explore the effects of the female-friendly design features on exhibits with greater initial discrepancies. In the present study, Kid Statements included self-talk, talk aimed at other children, and talk aimed at their caregivers. Researchers may want to explore the target of the girls' utterances to determine whether this narrows the effect and helps explain the dynamic between the exhibit, the girls, and their caregivers at the Non-Featured and Female-Friendly Featured versions of this or other exhibits. It is also possible that only certain types of girls' statements are affected by the inclusion of female-friendly features. Therefore, future researchers may also want to explore the type of children's statements to determine the effect of the female-friendly features on girls. These results, while encouraging for future research, do not currently support the theoretical framework that children's social interactions would increase and elicit more Contextualist-based scaffolding at the Female-Friendly Featured exhibit.

The results for Kids' Questions were inconclusive; there were no significant differences in the number of Kids' Questions based on the inclusion of the female-friendly design features for girls or for boys. These results are similar to those from Crowley's study (2001), in which the number of girls' questions did not differ significantly between the two versions of the exhibit. To further understand this result, an additional analysis was conducted to explore whether there were gender differences at the Non-Featured exhibit that would warrant (post-hoc) the need to enhance the number of Questions asked by girls to gain more gender equity at this exhibit and no significant gender difference was detected. While these results are inconclusive, the small effect sizes and their consistency with previous research suggest that this may be a less-fruitful area for further investigation. If future research in this area is conducted, it will be important to choose an exhibit where an initial discrepancy does exist. Kids' Social Interactions with their caregivers do not help explain the Engagement results; therefore, we turn now to caregivers' Social Interactions with their children.

**Caregivers' Social Interactions with Girls and Boys**

The results regarding the effects of the female-friendly features on caregivers' Social Interactions with their children were inconclusive. Specifically, this study did not detect any significant effects of the female-friendly features on the number or type of explanations (Meaningful or Directions/Procedure) caregivers provided, or the number of Questions caregivers asked, their girls or boys. There are several potential explanations for these results. The non-significant effects could be due to a lack of effect, or due to insufficient power or the particular sample included in the study. Alternatively, the need to increase caregivers' Meaningful Explanations and Questions to their girls stemmed from an expectation of initial inequity in caregivers' Social Interactions (based on previous work by Crowley, 2001), and a follow-up analysis was unable to detect any significant gender differences in caregiver Explanations or Questions at the Non-Featured version of the exhibit. Future researchers will want to identify exhibits where gender discrepancies in caregivers' Social Interactions exist prior to attempting to mitigate those differences by adding female-friendly design features. Theoretically, these results are surprising because the female-friendly version of the exhibit was designed to be more social, yet it did not encourage adults to provide more, or more meaningful, scaffolding for their children as Contextualist theory would suggest (Miller, 1993; Vygotsky, 1978). Future research should also explore other types of caregiver scaffolding such as modeling, joint participation, and encouragement (Miller, 1993).

Unlike the current study, Crowley found evidence of a gender gap in adults' verbal social interactions, and that adding a female user to the graphics enhanced caregivers' informative talk with their girls (K. Crowley, personal communication, October, 22, 2007; Crowley, 2001). While the current results may be counter to Crowley's for several reasons, it is possible that the focal participants' age or the museum context explain the different findings. The ages of the participants differed across the two studies; Crowley's study focused on children ages 1-8 with a median age of 4 or 5 (2001), while the current study focused on older children aged 8-12. Another study that was similarly unable to replicate Crowley's findings also had an older age group (median age of 7, range not reported; Borun, 1999). It is possible that such differences in social interactions are not seen with caregivers of older children. Another way that this study and Borun's differed from Crowley's was in the type of museum; the former were both conducted in science museums, while Crowley's was conducted in a children's museum. It is possible that parents who bring their girls to a science museum are less likely to show gender bias when interacting with their children at exhibits. The fact that caregivers' Social Interactions were similar for girls and boys even at the Non-Featured exhibit provides some support for this interpretation. These results could be partially explained by developmental systems and Contextualist theories, both of which emphasize the importance of context in understanding human behavior (Bronfenbrenner, 1998; Miller, 1993).

Future research would benefit from data collection with children of varying ages, on multiple exhibits and in multiple institutions (e.g., children's museums, zoos, aquaria, and science centers) to understand the applicability and generalizability of both the gender gap issues and female-friendly design features. It would also be interesting to determine the caregivers' drives for any differences between institutions, as well as design features to mitigate those differences. Researchers may also want to explore other types of caregiver talk, such as encouraging or discouraging remarks. However, the current study did explore Questions asked by caregivers, and again did not detect any significant differences at the two versions of the exhibit for girls or for boys. While these results may be due to the sample or sample size, they are also indicative of a lack of initial differences between the number of Questions asked of boys and the number asked of girls at the Non-Featured exhibit, $t(150) = -.84$, $p = .41$. While the results regarding caregivers' Questions are inconclusive, there is reason to believe that future research efforts in this area may be less fruitful than others. For example, the results related to caregivers' Questions are consistent with findings from a similar study that looked at the effects of enhancing exhibits with *family*-friendly design features (across multiple types of institutions) and did not find significant differences in the number of questions asked in the science museum setting (Borun, et al., 1998).

It is also important to acknowledge that the results of this study may have differed had other family dynamic variables been included. For example, a caregiver's gender, and mixed-gender versus same-sex dyads may also influence the Social Interactions of boys and girls and their caregivers at exhibits. Researchers have found that fathers are more likely to use causal explanations and scientific vocabulary with sons than with daughters (Tenenbaum & Leaper, 2003). However, the same researchers have found that mothers also use a higher proportion of scientific process talk with boys than girls (Tenebaum, Snow, Roach, & Kurland, 2005). Additionally, family dynamics for groups with more than one child or caregiver likely play a role in Social Interactions at exhibits. It was not possible to include these variables in the current study and still obtain sample sizes to achieve sufficient power; however, future researchers may want to explore how visitor group makeup affects girls' Engagement and Social Interactions.

It is surprising that none of the Social Interaction data helps explain the Duration results for girls with caregivers. Clearly, something more is happening for girls and their caregivers while they are spending significantly more time at the Female-Friendly Featured exhibit. Future research should consider other possible verbal and physical Social Interactions to try to elucidate the differential experiences at the two types of exhibits.

**Strengths and Limitations**

The applied nature and context of this research provides strong ecological validity to the results and allows for its immediate applicability within a science museum setting. The systems approach to the research encouraged a deep investigation into dynamic processes at the exhibit, especially in regard to the effects of the female-friendly design features on caregiver-child social interactions, and could be even more pronounced in the next round of related research studies. The systemic theoretical approach to museum studies is common among developmental psychologists who study museum learning (e.g., Crowley, 2000; Leinhardt & Crowley, 1998; Rogoff, 1990; Siegel, Esterly, Callanan, Wright, & Navarro, 2007), but less so among museum researchers from other fields; this systems theoretical stance could influence and inform future museum studies and their approach to exploring dynamic learning processes and interactions in visitor studies. Ecologically valid studies conducted in contexts as dynamic as museums, as with the current research design, often entail several trade-offs with other types of validity. Threats to internal and external validity were acknowledged early on in the study and whenever possible, measures were taken to mitigate those threats and their implications. Below, those threats and measures, along with two additional weaknesses of the study, power and assumptions, are discussed.

**Threats to internal validity.** The three main threats to the internal validity are the between-subjects design, coding scheme accuracy, and treatment integrity. The between-subjects design opens the significant Attraction and Duration findings to the possibility that the results could be due to fundamental and unrelated differences between the groups at each version of the exhibit rather than the inclusion of the design features. Similarly, the between subjects design may have made it difficult to detect any Social Interaction effects if the visitor groups were fundamentally different from one another. I attempted to mitigate the potential for differences between groups by collecting data over several weeks, rather than at a single moment in time, and approximating randomization by counterbalancing across days of the week (Eddington, 1989). Future researchers may want to incorporate a within-subjects component by including a second pair of exhibits and placing the Non-Featured from one pair in the room with the Female-Friendly Featured from the other pair (and vise-versa). Alternatively, researchers may want to employ a control exhibit to rule out differences between the groups unrelated to the female-friendly design features.

It is possible that the coding scheme did not accurately capture visitors' number or type of Statements, Questions, or Explanations. The current research borrowed from the coding schemes used by Crowley (2000, 2001) and Gutwill (Humphrey & Gutwill, 2005) to build on previous work. The adaptation of existing coding schemes, and the use of a pair coder in making those adaptations, allows for greater construct validity by ensuring that other researchers find face validity to the codes. Potential effects of the coding scheme definitions were further explored by analyzing the data using more tightly defined Kid Statements and Adult Informative Talk; results did not differ, lending additional construct validity to the scheme. Additionally, the high reliability scores suggest that it is unlikely that coders inaccurately applied the scheme. However, poor audio quality, as evidenced by the lower transcription percentage agreement, may have made it more difficult for the coders to accurately apply the scheme. Replications with the same coding scheme or validation with other measures or schemes should be considered in the future.

Another of the potential limits to the coding scheme involves the coding implementation. For example, adult talk was only coded if it was clearly aimed at the focal child; however, adults may intend for talk aimed at any of their children to be aimed at the entire family and therefore, not reiterate any Adult Informative Talk or Questions to the focal child. The coding scheme did not consider Adult Informative Talk aimed at other children in the group. Additionally, adults may have begun using the exhibit with another child in the visitor group, and even though they were using the exhibit alongside the focal child, they may not have been using it directly with the focal child; the coding scheme did not account for this type of interaction. The Social Interaction results may have differed if the coding scheme had been applied differently.

In regard to treatment integrity, it is important to acknowledge that the particular changes implemented at the Female-Friendly Featured exhibit may not have been the best possible representations of the design features to reach the design goals. I attempted to reduce this limitation by creating a knowledgeable group of advisors to brainstorm and vet ideas and changes for the Female Friendly Featured version based on the design goals and features. I also conducted formative evaluations on design changes whenever possible to provide quality checks along the way and ensure that the changes addressed the design goals without inhibiting exhibit usability. However, it is still possible that different female-friendly design choices would have led to stronger Social Interaction results, or weaker Attraction and Duration results.

**Threats to external validity.** There are three observable threats to external validity: single exhibit and setting, multiple changes, and sample selection. Results may reflect this particular exhibit's (i.e., *Geometry in Motion's)* sensitivity to these design features rather than effects that can generalize to other exhibits. However, *Geometry in Motion* is not egregiously different from many other science center exhibits; it is similar in pedagogical approach and design to exhibits that support visitor's open-ended exploration (Humphrey & Gutwill, 2005). As with all social science research, replication across multiple exhibits will be vital in order to generalize the effects of incorporating female-friendly design features on visitor behavior.

The single venue for conducting this research may reduce the generalizability of the results. That is, results for a science museum may not generalize to other institutions such as zoos or children's museums. As mentioned earlier, this is one of the more likely reasons that this study (and Borun's study, 1999) was unable to replicate Crowley's (2001) findings from a children's museum.

This research was designed to implement multiple changes to the exhibit at once; results may not generalize to exhibits that incorporate only a few of the female-friendly design features. It is unfortunate that this study could not tease apart the effects of single design features. However, the findings do provide an overarching rationale for future research that can more closely investigate each feature's contribution to girls' engagement at exhibits.

Finally, the results may not generalize beyond the sample chosen for the study. For instance, the findings may not apply to children younger than 7 or older than 13. It is also important to note that the sample for the Duration and Social Interaction data are a special subset of people who have chosen to use the exhibit. While this is not ideal because those results solely apply to visitors who are already interested in the exhibit, the Attraction data supports the notion that the female-friendly features have a broader application.

**Power.** An important weakness in this study is the insufficient power to detect interaction effects in the Duration and Kid Statements results. It is possible that this was the case for other analyses, but for these specific variables the pattern of the results and the effect sizes for the simple effects more prominently suggest that an interaction may be present. However, the large variability in Duration and Kid Statements, along with the sample size chosen for the study, mitigated the possibility of detecting a significant interaction, leading to inconclusive results. Future studies will want to include greater standard errors when conducing power calculations to determine sample size.

**Assumptions.** One of the major weaknesses for the social interaction research questions in the study is the assumption of initial gender inequity. Based on prior research in a children's museum (Crowley, 2001), this study inaccurately assumed that there would be gender disparity in the number and type of caregivers' Explanations to their children. This study extended that assumption to the number of caregivers' Questions asked of their children. Further, based on a systems theoretical view of caregiver/child transactions (Bronfenbrenner, 1998), this study also somewhat inaccurately assumed that there would be gender disparity in the number of children's Statements and Questions at the exhibit. These assumptions led to predictions of gains for girls that may not have been warranted, given that there was not a significant gender gap in these social interactions at the Non-Featured version of the exhibit.

## Implications

This research has a variety of implications, the majority of which are in the realm of practical applications. However, there are also implications for future researchers, for the theoretical framework (or logic model), and for developmental theory. Below, the implications of the results for each research question are discussed in detail.

The engagement results have the strongest practical implications. The increased attraction and time spent for girls allows for immediate consideration of the female-friendly design goals in exhibit development processes. The findings suggest that incorporating many female-friendly design features leads to deeper engagement for girls, providing a strong argument for considering these features in future science exhibit development projects. It is encouraging that the study was unable to detect any unintended negative effects for boys, which would discourage developers from incorporating these features in future exhibits. However, such non-significant results are inconclusive and should not dissuade future museum research and design teams from continuing to check for unanticipated ill effects for boys when incorporating female-friendly design features. These findings should encourage future researchers to continue to explore the individual effects of each of the female-friendly design features, and to employ alternative research designs to help triangulate the findings.

The engagement results support the original theoretical framework. Science museums have the opportunity to address topics where inequity exists, such as physical science and geometry, in a manner that engages females and helps them elude and reinvent cultural gender norms (Ruble, et al., 2006). Furthermore, female-friendly exhibits may play a role in reducing the gender gap in museum attendance by engaging girls more fully and comfortably, thus enticing them to return to this and other science museums (Borun, 1999). If female-friendly exhibits help encourage science museum attendance for girls, increased attendance could distally affect the gender gaps in spatial reasoning and STEM careers (Hamilton, et al., 1995; Salmi, 2001, 2002).

The engagement analyses were not aimed at testing developmental theory, but one of the initial findings does support the Contextualist theoretical perspective on learning. It was found that girls and boys spend significantly more time at the exhibit when with a caregiver than when alone. One of the major indicators of learning within a museum is the amount of time spent (Falk, 1983). Therefore, this result supports the Contextualist idea that learning is more likely to occur when a more knowledgeable other (i.e., caregiver) is present.

While the results of this study did not provide a conceptual replication of Crowley's studies of parental explanations (Crowley et al., 2001; K. Crowley, personal communication, October 27, 2007), they do encourage further explorations into the dynamics between caregivers and children at science museum exhibits, especially when female friendly features are incorporated. Contextualists posit that children actively participate in explorations that engage their social partners (Miller, 1993). Unfortunately, the Kid Statement results for this study were unable to provide evidence to support this notion; however, this appears to be a promising area for future study. This study was unable to identify any significant impact of the female-friendly design features on caregiver social interactions with their children either. Therefore, the study was also unable to provide support for theoretical framework in regards to the prediction that a carefully designed exhibit could provide caregivers with entry points to scaffold their children (Miller, 1993). The lack of gender differences at the Non-Featured version of the exhibit may be indicative of a contextual effect; systems theorists and Contextualists, such as Bronfebrenner (1998), would encourage future researchers to learn more about the gender effects across different informal learning contexts. One important implication for future researchers is the recognized need to choose exhibits where gender gaps in social interaction exist in order to study the effects of incorporating female-friendly design features on gender equity.

In summary, the first research question was supported: girls' Engagement at the exhibit did depend on whether the female-friendly design features were included. That is, girls were more attracted to, and spent more time at, the Female-Friendly Featured version of the exhibit. This study did not provide support for the second research question, which hypothesized that girls' quality of Social Interactions with caregivers at the exhibit would depend on whether the female-friendly design features were included. However, the results for Kid Statements suggest that this is a promising area for further investigation. Finally, the third research question, which investigated whether caregivers' Social Interactions with their girls at the exhibit would depend on whether the female-friendly design features were included, was not supported.

In conclusion, this research provides support for future practical and theoretical endeavors. This study has identified a variety of female-friendly design features and their promise for making exhibits more engaging for girls. It is my hope that this research will provide support for an NSF grant to more systematically investigate the results of each female-friendly design feature across multiple exhibits and institutions. There are many favorable routes for future research, and the current data set may even allow for additional explorations of caregiver/child interactions to help explain the Engagement results.

**References**


Abelson, R. P. (1996). Vulnerability of contrast tests to simpler interactions: An addendum to Rosnow and Rosenthal. *Psychological Science, 7*(4), 242-246.

Anderson, D., Zhang, Z., Chatterjee, S., Robin, R., & Aldrich, P. (2005). Punjab students' perceptions of science topics: Challenging outcomes from a front-end study. *Visitor Studies Today!, 8*(1), 8-12.

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*(5/6), 327-343.

Baenninger, M., & Newcombe, N. (1995). Environmental input into the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences, 7*(4), 363-379.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (Second ed.). Cambridge, UK: Cambridge University Press.

Baker, & Leary. (1995). Letting girls speak out about science. *Journal of Research in Science Teaching, 32*, 3-27.

Baranowski, M., & Delorey, J. (2007). Because dreams need doing: New messages for enhancing public understanding of engineering (pp. 46): National Academy of Engineering; BBMG; and Global Strategy Group.

Battista, M. T., Clements, D. H., Arnoff, J., & Van Auken Borrow, K. a. C. (1998). Students' Spatial Structuring of 2D Arrays of Squares. *Journal for Research in Mathematics Education, 29*(5), 503-532.

Ben-Chaim, D., G., Lappan, G., & Houang, R. T. (1985). Visualizing Rectangular Solids Made of Small Cubes: Analyzing and Effecting Students' Performance. *Educational Studies in Mathematics, 16*, 389-409.

Berg, A., & Lie, M. (1995). Feminism and constructivism - do artifacts have gender? *Science Technology & Human Values, 20*(3), 332-351.

Bjorklund, D. F., & Douglas Brown, R. (1998). Physical play and cognitive development: Integrating activity, cognition, and education. *Child Development, 69*(3), 604-606.

Borun, M. (1999). Gender roles in science museum learning. *Visitor Studies Today!, 3*(3), 11-14.

Borun, M., & Dritsas, J. (1997). Developing family-friendly exhibits. *Curator, 40*(3), 178-192.

Borun, M., Dritsas, J., Johnson, J., Peter, N. E., Fadigan, K., Jangaard, K. (1998). *Family Learning in Museums: The PISEC Perspective*. Washington D.C.: National Science Foundation.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In R. M. Lerner (Ed.), *Handbook of child psychology: Theoretical models of human development* (Vol. 1, pp. 993-1028). New York: Wiley.

Bryden, M. P. (1982). Sex differences in laterality *Laterality: Functional asymmetry in the intact brain*. New York: Academic Press.

Burton, L. A., Henninger, D., & Hafetz, J. (2005). Gender differences in relations of mental rotation, verbal fluency, and SAT scores to finger length ratios as hormonal indexes. *Developmental Neuropsychology, 28*(1), 493-505.

Casey, M. B. (1996). Understanding individual differences in spatial ability within females: A nature/nurture interactionist framework. *Developmental Review, 16*, 241-260.

Casey, M. B., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction, 26*, 269-309.

Casey, M. B., Erkut, S., Ceder, I., & Mercer Young, J. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29*, 29-48.

Casey, M. B., L., N. R., & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology, 35*(5), 1237-1247.

Cassell, J., & Jenkins, H. (2000). Chess for girls? Feminism and computer games. In J. Cassell & H. Jenkins (Eds.), *From Barbie to Mortal Kombat: Gender and computer games*. Cambridge, MA: MIT Press.

Cavallo, D., Sipitakiat, A., Basu, A., Bryant, B., Welti-Santos, L., Maloney, J. (2004). *RoBallet: Exploring Learning through Expression in the Arts through Constructing in a Technologically Immersive Environment*. Paper presented at the International Conference of the Learning Sciences, California, USA.

Ceci, S. J., & Williams, W. M. (2007). *Why Aren't More Women in Science? Top Researchers Debate the Evidence*. Washington DC: American Psychological Association.

Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal, 98*(2), 171-187.

Coley, R. (2001). Differences in the gender gap: Comparisons across racial/ethnic groups in education and work. Retrieved September 1, 2008

Costa, P., Terracciano, A., & McCrae, R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*(2), 322-331.

Crowley, K. (2000). Parent explanations during museum visits: Gender differences in how children hear informal science. *Visitor Studies Today!, 3*(3), 21-28.

Crowley, K. (2001). Kevin Crowley talk: Just in time explanatory support from Power Girl. Pittsburgh Scientific Reasoning Supergroup. http://www.lrdc.pitt.edu/supergroup/

Crowley, K., Callanan, M., Tenenbaum, H., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. *Psychological Science, 12*(3), 258-261.

Deaux, K. (1985). Sex and gender. *Annual Reviews of Psychology, 36*, 49-81.

Diamond, J. (1994). Sex differences in science museums: A review. *Curator, 37*(1), 17-24.

Dick, W., & Carey, L. (1996). *The Systematic design of instruction*. New York: Harper Collins Publishing.

Eddington, E. S. (1989). Approximate randomization tests. *Journal of Psychology, 92*, 143-149.

Else-Quest, N. M., Linn, M. C., & Shibley Hyde, J. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127.

Erickson, G., & Farkas, S. (1991). Prior experience and gender differences in science achievement. *The Alberta Journal of Educational Research, 37*(3), 225-239.

Exploratorium. (2008). Fact sheet: 2007-2008 (pp. 1): Exploratorium.

Falk, J. H. (1983). Time and behavior as predictors of learning. *Science Education, 67*(2), 267-276.

Ferrini-Mundy, J. (1987). Spatial training for calculus students: Sex differences in achievement and in visualization ability. *Journal for Research in Mathematics Education, 18*(2), 126-140.

Finn, K. (2007, July 17, 2007). [Have you noticed gender differences on the floor at the Exploratorium?].

Flaherty, M. (2005). Gender differences in mental rotation ability in three cultures: Ireland, Ecuador and Japan. *Psychologia, 48*, 31-38.

Ford, A. (1999). Bull's-eye diagrams *Modeling the environment: An introduction to system dynamics models of environmental systems*. Washington, DC: Island Press.

Ford, D. J., Brickhouse, N., Lottero-Perdue, P., & Kittleson, J. (2006). Elementary girls' science reading at home. *Science Education, 90*, 270-288.

Froschl, M., Sprung, B., Archer, E., & Fancsali, C. (2003). Science, gender, and afterschool: A research-action agenda. Retrieved June 22, 2007, from http://www.afterschool.org./sga/content/strategies.cfm

Gallagher, L., & Michalchik, V. (2007). Assessing youth impact of the computer clubhouse network: May 2007 Youth Impact Survey Administration Retrieved October 8, 2007

Greenfield, T. A. (1995). Sex differences in science museum exhibit attraction. *Journal of Research in Science Teaching, 32*(9), 925-938.

Gutwill, J. (2003). Gaining visitor consent for research II: Improving the posted-sign method. *Curator, 46*(2), 228-235.

Hall, J. A., & Murphy, N. A. (1996). Recall of nonverbal cues: Exploring a new definition of interpersonal sensitivity. *Journal of Nonverbal Behavior 30*, 141-155.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psycholbiosocial model to understand cognitive sex differences. *Brain and Cognition, 45*, 392-414.

Hamilton, L. S., Nussbaum, E. M., Kupermintz, H., Kerkhoven, J. I. M., & Snow, R. (1995). Enhancing the validity and usefulness of large-scale educational assessments: II. NELS:88 science achievement. *American Educational Research*

*Journal, 32*(3), 555-581. doi: http://www.jstor.org.proxy.lib.pdx.edu/view/00028312/ap040128/04a00060/0

Hanna, G. (1990). Mathematics achievement of boys and girls: An international perspective. *Ontario Mathematics Gazette: Ontario Association of Mathematics Education, 28*(3), 28-32.

Hill, O. W., Pettus, W. C., & Hedin, B. A. (1990). Three studies of factors affecting the attitudes of blacks and females toward the pursuit of science and science-related careers. *Journal of Research in Science Teaching, 27*(4), 289-314.

Howell, D. C. (2002). *Statistical Methods for Psychology* (5th edition ed.). Pacific Grove, CA: Duxbury.

Humphrey, T., & Gutwill, J. (2005). *Fostering Active Prolonged Engagement: The Art of Creating APE Exhibits* San Francisco: Left Coast Press.

Jacobs, J. E., Finken, L. L., Griffin, N. L., & Wright, J. D. (1998). The career plans of science talented rural adolescent girls. *American Educational Research Journal, 35*, 681–704.

Jaschik, S. (2007). 'Why aren't more women in science?'. *www.insidehighered.com/news/2007/01/03/women*. Retrieved from doi:www.insidehighered.com/news/2007/01/03/women

Jenkins, H. (2001). From Barbie to Mortal Combat: Further reflections; http://culturalpolicy.uchicago.edu/conf2001/papers/jenkins.html Retrieved June 22, 2007

Jones, G., Howe, A., & Rua, M. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education, 84*, 180-192.

Kekelis, L., Heber, E., & Countryman, J. (2005). A bridge to technology: Designing a program that attracts girls. *ASTC Dimensions, May/June*, 3-4.

Kimball, M. M. (2007). Adding Gender to the Mix: A Commentary on `Toward a Redefinition of Sex and Gender' *Feminism & Psychology, 17*, 453-459.

Kimura, D. (1999). *Sex and Cognition*. Cambridge, MA: MIT Press.

Koke, J. (2005). Barriers to choice: How adolescent girls view science careers. *ASTC Dimensions, May/June*, 5.

Kozulin, A. (Ed.). (1986). *Thought and Language*. Cambridge, MA: MIT Press.

Kremer, K. B., & Mullins, G. W. (1992). Children's gender behavior at science museum exhibits. *Curator, 35*(1), 39-48.

Kutnick, P., & Kington, A. (2005). Children's friendships and learning in school: Cognitive enhancement through social interaction. *British Journal of Educational Psychology, 75*, 521-538.

Lake, C. (2005). Design psychology: Gender and user experience. An interview with Gloria Moss. *http://designpsych.weblog.glam.ac.uk/2005/10/13/gender-and-user-experience* Retrieved October 10, 2009

Lee, V. E., & Burkam, D. T. (1996). Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis. *Science Education, 80*(6), 613-650.

Leinhardt, G., & Crowley, K. (1998). Museum learning as conversational elaboration: A proposal to capture, code, and analyze talk in museums. (Museum Learning Collaborative Technical Report # MLC-01). Pittsburgh, PA: Learning Research & Development Center at University of Pittsburgh.

Lendaris, G. (1986). On systemsness and the problem solver: Tutorial comments. *Systems, Man, and Cybernetics, 16*(4), 604-610.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.

Lubinski, D., & Persson Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on psychological science, 1*(4), 316-345. doi: http://pps.sagepub.com/content/1/4/316.short

Maher, K. (2005). Building girls: The science of home improvement. *ASTC Dimensions, May/June*, 6-7.

Mann, V. A., Sasanuma, S., Sakuma, N., & Masaki, S. (1990). Sex differences in cognitive abilities: A cross-cultural perspective. *Neuropsychologia, 28*(2), 1063-1077.

Martin, C. L., & Ruble, D. N. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current directions in psychological science:, 13*(2), 67-70.

McCreedy, D. (2005). Engaging adults as advocates. *Curator, 48*(2), 158-176.

Medlock, M. (2005). The Rapid Iterative Test and Evaluation method: Better products in less time. In R. G. Bias & D. J. Mayhew (Eds.), *Cost-justifying usability: An update for an Internet age* (Second Edition ed., pp. 489-517). San Francisco, CA: Morgan Kaufmann Publishers.

Medlock, M., Wixon, D., Terrano, M., Romero, R., & Fulton, B. (2002). *Using the RITE method to improve products: A definition and a case study*. Usability Professionals Association, ( URL http://download.microsoft.com/download/5/c/c/5cc406a0-0f87-4b94-bf80-dbc707db4fe1/mgsut_MWTRF02.doc.). Cited in: Cost-justifying usability: An update for an Internet age (edited by Bias and Mayhew).

Miles, M., & Humberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (Second ed.). Thousand Oaks: Sage Publications.

Milgram, D. (2005). Gender differences in learning style specific to science, technology, engineering and math Retrieved October 5, 2007

Miller, P. H. (1993). *Theories of developmental psychology* (3rd ed.). New York, NY: W. H. Freeman and Company.

Mintz, J. (2007). New software design may help women users; http://www.boston.com/business/globe/articles/2007/09/24/new_software_design_may_help_women_users/. *Boston.com* Retrieved October 5, 2007

Morgan, C., Isaac, J. D., & Sansone, C. (2001). The Role of Interest in Understanding the Career Choices of Female and Male College Students. *Sex Roles, 44*(5/6).

Moss, G., Gunn, R., & Heller, J. (2006). Some men like it black, some women like it pink: Consumer implications of differences in male and female website design. *Journal of Consumer Behavior, 5*, 328-341.
Moss, G., Gunn, R., & Kubacki, K. (2007). Successes and failures of the mirroring principle: the case of angling and beauty websites. . *International Journal of Consumer Studies, 31*, 248-257.
National Center for Education Statistics. (2003). *Trends in International mathematics and science study (TIMSS)*. Available at http://nces.ed.gov/timss/results03.asp.
National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: National Academies Press.
National Science Board. (2008). Science and Engineering Indicators 2008. Vol. 1. doi:http://www.nsf.gov/statistics/seind08/
National Science Foundation. (2002a). Gender differences in the careers of academic scientists and engineers (Vol. NSF 04-323). Arlington, VA.
National Science Foundation. (2002b). NSF's Program for Gender Equity in Science, Technology, Engineering, and Mathematics: A brief retrospective 1993-2001 (Vol. NSF 02-107.). Arlington, VA.
National Science Foundation. (2003). *New formulas for America's workforce: Girls in science and engineering*. Arlington, VA: NSF 03-207.
Nesbit, J. C., & Hadwin, A. F. (2006). Methodological issues in educational psychology. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (Second ed., pp. 825-847): Routledge.
OMSI Evaluation & Visitor Studies Department. (2008). *Access Algebra Front End Report Findings*. OMSI. Portland, Oregon.
Pallrand, G., & Seeber, F. (1984). Spatial Ability and Achievement in Introductory Physics. *Journal of Research in Science Teaching, 21*, 507-516.
Peterson, P., & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal, 22*(3), 309-335.
Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology, 81*(4), 639-646.
Piaget, J., & Inhelder, B. (1967) *The child's conception of space* (pp. 375-418). New York: Norton.
Pontius, A. A. (1991). Cross-evolutionary spatial representation in stone-age ecology. *Behavioral and Brain Sciences, 14*(3), 522-523.
Quaiser-Pohl, C., & Lehmann, W. (2002). Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology, 72*, 245-260.
Ramey-Gassert, L. (1997). Learning science beyond the classroom. *The Elementary School Journal, 97*(4), 433-450.

Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond Black Boxes: Bringing transparency and aesthetics back to scientific investigation. In the Journal of the Learning Sciences. *Journal of the Learning Sciences* Retrieved October 8, 2007

Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York, NY: Oxford University Press.

Roscoe, R. (2008). Analyzing qualitative verbal data. *www.pitt.edu/~roscoe/verbalanalysis.pdf* Retrieved November, 20, 2008

Rosenthal, R., & Rosnow, R. L. (2008). *Essentials of behavioral research: Methods and data analysis* (3rd ed.). Boston: McGraw Hill.

Rosnow, R. L., & Rosenthal, R. (1996). Contrasts and interactions redux: Five easy pieces. *Psychological Science, 7*(4), 253-257.

Rosser, S. V. (1991). *Female friendly science: Applying women' s studies methods and theories to attract students*. New York: Teacher's College Press.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender Development. In W. Damon, R. M. Lerner & N. Eisenberg (Eds.), *Handbook of Child Psychology: Social, Emotional, and Personality Development* (6th ed., Vol. 3, pp. 858-932). Hoboken, NJ: Wiley & Sons, Inc.

Salinger, S., Plonka, L., & Prechelt, L. (2008). A coding scheme development methodology using grounded theory for qualitative analysis of pair programming. *Human Technology, 4*(1), 9-25.

Salmi, H. (2001). *Public understanding of science: universities and science centres*. Paper presented at the Management of University Museums; Education and skills; OECD, Paris.

Salmi, H. (2002). Factors affecting students' choice of academic studies: the motivation created by informal learning (Publication no. http://www.heureka.fi/portal/englanti/about_heureka/research/factors_affecting_students%60_choice_of_academic_studies__the_motivation_created_by_informal_learning._/). Retrieved October 10, 2008, from Heureka: The Finnish Science Center

Schneider, B., & Cheslock, N. (2003). Measuring results: Gaining insight on behavior change strategies and evaluation methods from environmental education, museum, health, and social marketing programs. San Francisco: Coevolution Institute.

Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Shibley Hyde, J. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

Shibley Hyde, J. (2008). Gender similarities characterize math performance. *Science, 321*, 494-495. doi: sciencemag.org

Shibley Hyde, J., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139-155.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin, 86*(2), 420-428. doi: http://www3.uta.edu/faculty/ricard/COED/Shrout%20&%20Fleiss%20(1979)%20Intraclass%20correlations.pdf

Siegel, D. R., Esterly, J., Callanan, M. A., Wright, R., & Navarro, R. (2007). Conversations about science across activities in Mexican-descent families. *International Journal of Science Education, 29*(12), 1447 - 1466.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics. *American Psychologist, 60*(9), 950-958.

Subrahmanyam, K., & Greenfield, P. M. (2000). Computer games for girls: What makes them play? In J. Cassell & H. Jenkins (Eds.), *From Barbie to Mortal Kombat: Gender and computer games*. Cambridge, MA: MIT Press.

Super, D. E., & Bachrach, P. B. (1957). *Scientific careers and vocational development theory*. . New York, NY: Bureau of Publications, Teachers College, Columbia University.

Sylvan, E. (2005). Integrating aesthetic, engineering, and scientific understanding in a hands-on design activity  Retrieved October 8, 2007

Tarte, L. (1990). Spatial Orientation Skill and Mathematical Problem Solving. *Journal for Research in Mathematics Education 21*(3), 216-229.

Taylor, D. (2002). Gender differences in informal education environments: A review of the literature on gender and learning in science museums. *The Informal Learning Review, Jan. - Feb.*, 6-7.

Taylor, D. (2005). Social science: Observing women and girls in the museum. *ASTC Dimensions, May/June*, 11-12.

Taylor, D. (2006). Girls, boys, moms, and dads: Learning about their different needs in science museums. In C. Yao (Ed.), *Handbook for Small Science Centers* (pp. 310). Lanham, MD: Altamira.

Tenebaum, H. R., Snow, C. E., Roach, K. A., & Kurland, B. (2005). Talking and reading science: Longitudinal data on sex differences in mother-child conversations in low-income families. *Applied Developmental Psychology, 26*, 1-19.

Tenenbaum, & Leaper. (2003). Parent-child conversations about science: Socialization of gender inequities. *Developmental Psychology, 39*, 34-47.

The American Heritage® Dictionary of the English Language. (retrieved September 07, 2008, http://dictionary.reference.com/browse/linkage). Linkage definition

Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs, 2*.

Tracy, D. (1987). Toys, Spatial Ability, and Science and Mathematics Achievement: Are they Related? . *Sex Roles, 17*(3/4), 115-138.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599-604.

Verheyden, P. (2003). The great sexperiment. *ECSITE Newsletter, 54*(Spring), 10-11.

Visser, P., Krosnick, J. A., & Lavrakas, P. J. (2000). Survey Research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 223-252). Cambridge, UK: Cambridge University Press.

Voyer, D. (1996). On the magnitude of laterality effects and sex differences in functional lateralities. *Laterality, 1*(1), 51-83.

Voyer, D., & Bryden, M. P. (1990). Gender, level of spatial ability, and lateralization of mental rotation. . *Brain & Cognition, 13*, 18-29.

Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles, 43*(11/12), 891-915.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitde of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cole, M., John-Steiner, V., Scribner, S., Souberman, E. (Eds.) Cambridge, MA: Harvard University Press.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817-835. doi: http://www.vanderbilt.edu/Peabody/SMPY/Wai2009SpatialAbility.pdf

West, S. G., Biesanz, J. C., & Pitts, S. (2000). Causal inference and generalization in field settings. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 40-84). Cambridge, UK: Cambridge University Press.

White, P. E. (1992). Women and minorities in science and engineering: An update (pp. xi, 27-60): National Science Foundation.

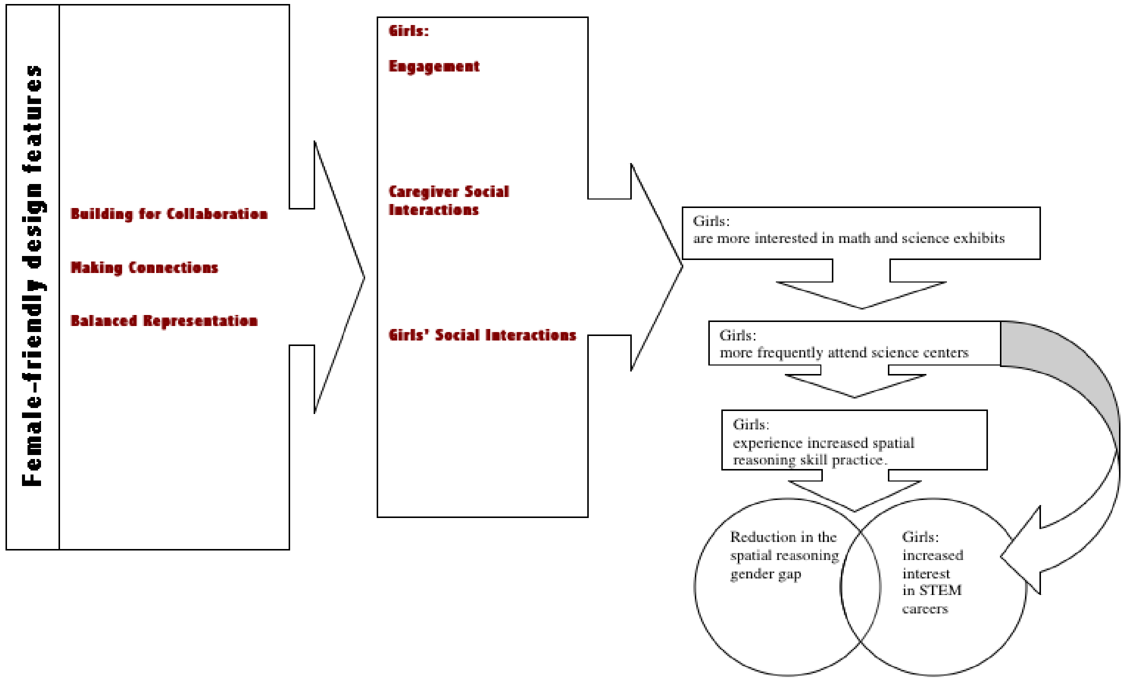
Wilde, M., & Urhahne, D. (2008). Museum learning: a study of motivation and learning achievement in educational research. *Journal of Borderland Education, 42*.

Witkin, H. A., Dyk, R. B., & Faterson, H. F. (1962). *Psychological differentiation*. New York: Wiley.

# Appendix A

Female-friendly exhibits: Logic model.

**Goal: Reduction in the spatial reasoning gender gap, and increased interest in STEM careers for females.**



Appendix B. Advisory team brainstorming notes and changes to exhibit.

| Design Goal | Area for improve-ment | Why need to address this area | **Possible Solutions** | Cons? | Addressing ? |
|---|---|---|---|---|---|
| Social Interaction & Collabora-tion | Allow for visitors to connect their linkages/ work individually and together in collaboration | Too little space to work together | **Larger tables (suggested rectangular)** | Not sure how this change would affect behavior. Really needed? People will still run off the table. Different from comparison exhibit… $$$ | Yes: We inset two of the circular tables adjacent to one another into a rectangular background—a compromise. |
| | | Hard to build across tables at different heights. | **Single height for each table.** | May be a less comfortable height for some. | Yes. See above. |
| | More collaboration and girls having more ownership of the area. | | Two-bum bench | Doesn't feel like a sitting activity … May prototype with and without bench | No: Changes to the storage system made the reach impossible with benches. |
| | | "knowing what I'd be in for" (e.g., a title like: Move Me to the Moon). | **Need a big colorful sign that is as noticeable as the mess of linkages on top of the tables.** | | Yes: Hired an external graphics designer to create a big colorful label that provided |

| Design Goal | Area for improve-ment | Why need to address this area | **Possible Solutions** | Cons? | Addressing? |
|---|---|---|---|---|---|
| | | | | | easier understanding of the exhibit upon approach. |
| | | Encourage collaboration | Label copy that suggests collaboration | Enough space? | No, focused on getting started. |
| Connecting to Social/Community & Providing Context | Needs a goal and or motivator.<br><br>Needs a storyline or opportunity for storyline | Need an activity that inspires building. What is the goal and its motivation? Creative goal: Problem to solve, story to tell or thing to reproduce. "why am I doing this? Complete a scene; help someone" Goal: "Make something happen! Having the linkage be the goal doesn't work when you can't build anything interesting and complicated yet. I want a reward that's bigger than my first dumb linkage" | **Puppets: Create whimsical things with linkages. Have Mr. Potato Head-like options (hands, eyes, feet: challenge, make them push each other, wave, clap, kick something; hammer hits a bell).**<br><br>**Pieces with top and bottom connection possible: claws, paws, hands, feet (with shoes—chucks?), feather, wing, needle, hammer, tennis racket)**<br><br>**Pieces with bottom connection possible: googly eyes and bell, maybe tennis ball, soccer ball. May get more Explo buy-in than storyboard or game-board (or not). Opportunity for visitors to create their own narrative, while not providing an absolute story.** | Only enough time for a single prototyping session. | Yes: We are including several of these pieces. We tried and kept the hammer and bell, the hands, the feet and the googly eyes. We also tried but did not keep (due to visitor confusion) wings and feathers. Challenges include: waive the hands, kicking feet, and ring the bell. |

| Design Goal | Area for improve-ment | Why need to address this area | **Possible Solutions** | Cons? | Addressing ? |
|---|---|---|---|---|---|
| | | | **That is, the storyline may emerge.** | | |
| | | | **Challenge objects. Whimsical, gravity-based (if table is vertical), feedback. Make a lion's cage open, an elephant turn, a bell ring. Start→ end. Monkey→ Banana.** | No clear ideas yet, other than the bell. | Yes: We included a hammer and a bell. |
| Balanced Represent-ation | Storyline and aesthetics | Attraction and engagement | **Whimsy: Try color (and topic.) Will be using 3 colors. Color code size of pieces.** | | Yes: We incorporated four new colors (each rod length has its own color). Colors were selected by asking girl visitors in the age group to choose their favorites from the current exhibition color palette. |
| | Aesthetics | Too messy | **Better trays (5" deep, long and wide but not deep)** | Need a rectangular edge. | Yes: We added shallow trays to the bottom of the exhibit—along the edge. |
| | | | **Tilted table (easel tilt ~35 degrees)** | | Yes: We created a |

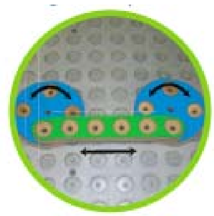
| Design Goal | Area for improve-ment | Why need to address this area | **Possible Solutions** | Cons? | Addressing ? |
|---|---|---|---|---|---|
| | | | | | table with an adjustable tilt and chose the best angle based on developer observations of visitor use. |
| | Aesthetics and environmental cue that it is for them | Pieces too heavy | **Shorten them (7, 5, 3 or 1, 3, 5?)** | | Yes: 2, 4 and 5. |
| | | Table too industrial | **Powdercoat? Beige, pale blue, white? Make the pieces pop.** | | Yes: We used a white laminated backdrop, and painted the table a very light warm grey (based on female developer suggestion). |
| | Label copy | Use wording that is comfortable to females, too. | **We changed the directive text (e.g., "Try This") to more relational text (e.g., "Here's the secret).** | | Yes: We included what we could while focusing on ease of use, too. |
| | | Include font that is attractive to females, too. | **Try colored and more rounded fonts in the label.** | | Yes. |

| Design Goal | Area for improve-ment | Why need to address this area | **Possible Solutions** | Cons? | Addressing ? |
|---|---|---|---|---|---|
| | Incorporate female voices | Several females felt:<br>1. Too messy,<br>2. Round is confusing and too small<br>3. Tiers are frustrating<br>4. Need a reason to build a linkage<br>5. Needs more color | **See above.** | | We are addressing most of these issues. 1. Yes: The tilt and new storage bin addressed the messiness.<br>2&3. Yes: The size and shape of each table didn't change, but they more readily allow for building across tables by insetting them adjacent and at the same height.<br>4. Yes: All of the additional pieces aim to provide a reason to build a linkage.<br>5. Yes: We are including at least 3 colors. |

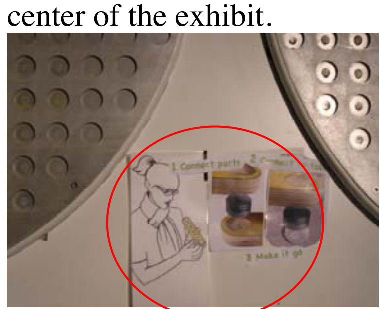
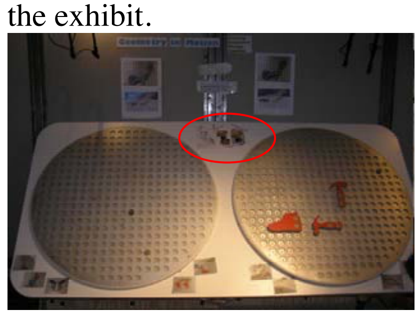
| Design Goal | Area for improve-ment | Why need to address this area | **Possible Solutions** | Cons? | Addressing ? |
|---|---|---|---|---|---|
| | Highlight female users | In alignment with Powergirl study—may enhance experience with parents and increase attraction | **Include a female user in the label** | | Yes. |

Appendix C. Changes based on iterative evaluation results.

| Area of change | Reason for change | Iterations | Final solution |
|---|---|---|---|
| Exhibit design: Tilt | To keep the surface area clean of unused parts while still allowing visitors to build linkages that didn't break due to gravity. | Steep (and mid-to-steep) tilt:<br><br>Shallow (and mid-to-shallow) tilt:<br> | Mid-to-shallow tilt. |
| Exhibit design: Number of components | Visitors were often confused by the iconography for the wing, feather and needle, and some appeared overwhelmed by the number of options. | 1. All components included.<br><br>2. Wing, feather, and needle removed.<br> | Removed wing, feather and needle. |

| Area of change | Reason for change | Iterations | Final solution |
|---|---|---|---|
| Label content: ‘Give it a push’ example | During observations it became clear that the ‘Give it a push’ example was the linkage most visitors’ first attempt. This drove us to find an example that ensured initial success (while avoiding a second view depicting all of the connection points). | **1.** Hammer- bell<br><br>2. Kicking knee<br><br>3. Train wheels<br> | Train wheels. |
| Label content: ‘What’s Going On’ text wording | To develop text that immediately communicated what to do and helped visitors when they became stuck. | 1.“Your sculptures can dance, kick, wave, or even clap hands with a friend across the table. You’re Making a Linkage. Want to change back-and-forth motion into round-and-round? That what linkages do, they connect moving parts and turn one kind of motion into another.”<br>2.“Do you and your friends want to clap the hands, ring the bell, or make the feet dance? Change where you connect these parts to each other or to the table and you can change the motion of your sculpture. It can go up-and-down, round-and-round, or back-and-forth.”<br>3.“If you change where you connect these parts to each other or to the table, you can change the motion of your sculpture.” | “If you change where you connect these parts to each other or to the table, you can change the motion of your sculpture.” |

| Area of change | Reason for change | Iterations | Final solution |
|---|---|---|---|
| Label content: Tagline | Iterating to find the best way to use relational language that is representative of females while still communicating the basics of what to do at the exhibit. | 1. Creating Moving Sculptures Together<br>2. Creating Moving Sculptures<br>3. You can create a sculpture that moves | You can create a sculpture that moves. |
| Label content: Number of Large examples | To determine whether visitors more often succeed in creating a linkage with two large examples (that each included a second view depicting all of the connection points) or a single large example (depicting the connection points). | 1. Two large examples (waving hand and kicking feet).<br><br>2. Single large example (waving hand).<br> | A single large example : waving hand. |
| Label content: Large example | To determine whether visitors were more successful at the waving hand or kicking foot. | 1 iteration: Tallied percentage of successful (rather than frustrated and incompleted) attempts when both kicking feet and waving hand examples were present. | Waving hand. |
| Label content and placement: | To determine where we could best place a | 1. mini examples placed around the large connecting circles | Mini example s |

| Area of change | Reason for change | Iterations | Final solution |
|---|---|---|---|
| Mini example photos | multitude of examples to help give people the idea that there are tons of cool linkages you can build. However, the mini examples, regardless of placement, introduced frustration and confusion. | <br>2. mini examples lining the front of the trough<br> | introduced too much confusion and were removed from the exhibit. |
| Label placement: 'How To' | Needed visitors to know what to do to get started (e.g., how to connect the parts) immediately when they approach the exhibit. | 1. 'how to' placed at the bottom center of the exhibit.<br><br>2. 'how to' placed at the top center of the exhibit.<br><br>3. 'how to' placed in the middle of the main label. | 'How To' placed in the middle of the top label. |

| Area of change | Reason for change | Iterations | Final solution |
|---|---|---|---|
| | |  | |
| Label placement: Placement of Large examples | Trying to determine the placement of the larger examples most conducive to successful completion of the waving hand. | 1. To the sides on the main label<br><br>2. To the sides flanking the exhibit.<br><br>3. To the left on the main label.<br><br>4. To the right on the main label.<br> | The single large example to the right of the main label. |

Appendix D. Coding Scheme

## Durations Coding Scheme

Eligible Focal Child:

1. Focal child must be between 47” and 62” in height
2. Child must touch the exhibit or stand feet planted and facing the exhibit for at least 3 seconds

**Duration is the length of time the visitor spends at the exhibit.**

*Start time*:

- When the focal child watches for a bit and then touches the exhibit, rewind the video to the moment when the focal child first plants both feet while facing the exhibit. Start from that point.
- For visitors who immediately touch the exhibit, start when they first touch the exhibit. If you can’t see, start the duration code when they plant both feet at the exhibit.
- For visitors who never touch the exhibit, only include them if they spend at least three seconds with feet planted facing the exhibit.

*Stop time:*

- Watch for the moment when the focal child leaves the exhibit. Rewind the video to the moment when the focal child **is no longer touching and their feet are no longer planted at the exhibit**.
    - o If the child’s feet are unplanted first, the code should end at just after the final touch.
    - o If the child is no longer touching the exhibit, but is still standing feet planted and facing the exhibit, the code should end just as they un-plant their feet.
    - o If you cannot see the final touch or moment when they move their feet, end the duration code when you are sure they are no longer touching the exhibit or standing feet planted.

*Return visits:*

- Always scan to the end of the clip to find any return visits. Create a second, third, etc… duration clip using the same start and stop rules as above. The data will later be combined to determine total time spent at the exhibit.
- To count as a return visit, the child must either spend at least three seconds with feet planted facing the exhibit, or touch the exhibit.

*See exemplary videos*.

- E2_SA_090523_vg7_EXEMP
- E2_TH_090618_s1_vg10_EXEMP
- E1_FR_090814_s1_vg2_EXEMP

**Social Interaction Coding Scheme**

Eligible visitors:

1. Groups of at least one adult and a focal child.
2. Focal child must be between 47” and 62” in height
3. Focal group must speak English while at the exhibit (entire time)
4. Focal group cannot be part of a camp/school/scout group (indicators include multiple people with nametags or uniforms).

Coding time: Coding should begin three seconds prior to the focal child’s 1st duration start time and continue until three seconds after the focal child’s last duration stop time.

Eligible visitor groups: Begin each coding session by listening through the visitor group interactions. If you are confused about who is speaking or what is being said for more than ~1/5 of the video, note the group in the missing data log and do not code. (if you are on the fence, do not code).

Determining who’s talking: It can be difficult to identify who in a group is speaking. Here are some suggestions that can help you in identifying the speaker:

- Play through the video or part of the video once to get a good feel for:
    - how the voices differ.
    - how the vocal tones differ.
        - slow down the tempo to get a good feel for vocal tones.
    - any names that are used (may help in deducing who is speaking or being spoken to).
- Kids often provide the following hints when they are talking:
    - They animate their utterances with body language (see E1_SU_090531_vg13, instances 4 and 5 for an example—the girl in pink is animating the sentence not the girl in white).
    - They physically punctuate their “here,” “there,” “this” type statements with obvious actions.

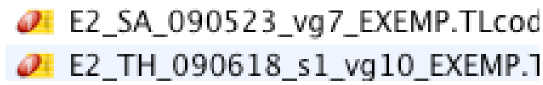
Eligible utterances: Visitor’s utterances are only eligible for coding when they are in the Sound Abatement area. Oftentimes an adult will leave the Sound Abatement (SA) area and converse with the child who is still at the exhibit; in this case, the child’s utterances made within the SA area can still be coded, but the adult’s comments from outside the SA area CANNOT be coded.

- If you absolutely cannot determine who is speaking, code the utterance as unintelligible.

Determining utterance meaning:

- It can be difficult to determine what is being said. Here are a couple of suggestions that we’ve found work well:
    - Listen a couple of times to what is said.
    - Be open to wording different from what you first hear.

    - Use the contextual conversation to determine meaning.
- Use only what the visitors say (immediately and in the surrounding conversational context) to determine how each utterance should be coded. Please do not guess at intent. Visitor intentions may seem very clear to each of us based on our own experiences; however, each of us will guess at different intentions because our prior experiences differ. Please take visitor utterances at face value, even though it is tempting to use your interpretation of the family dynamic to interpret what they "really mean." For example, a question should be coded as a question even if we think the intent is rhetorical.

Exemplary clips: 

& E1_FR_090814_s1_vg2_EXEMP

FOCAL CHILD

## Focal Child Utterances

**Codes for Focal Child Utterances:**

***KS*:** Kid Statement. Any statement produced by the focal child that has to do with the exhibit. This includes content-less utterances such as grunts, sound effects and singing - all of which can be interpreted as indications of a child's engagement or interest during the activity. Any remark or noise made by the focal child that has to do with the exhibit and is not clearly a question is a KS.
Note: If it is unclear what the child is saying, default to Kid Statement.
Note: If it is unclear if it is the focal child speaking, default to Unintelligible Talk.

***KQ:*** Kid Question. Any question asked by the focal child that has to do with the exhibit should be coded as KQ.

***Labels: The two labels for Kid Statements are Content-less and Notes. The labels are discussed in detail following the main code descriptions and examples.***

*Multiple kid utterances:* If a focal child says a string of utterances without pausing for 2 seconds, code them as a single utterance unless there are verbal utterances made by the adult(s) in between (this requires that they are not speaking at exactly the same moment, or over top of one another completely); or their multiple utterances include a combination of questions and statements. That is, quick changes in speaker, or from questions to statements and vice versa, should be broken out into separate codes. Strings that do not change from questions to statements or vice versa are considered an elaboration on the same utterance and should be coded as a single utterance.

*Deciding whether an utterance is to be coded as a question:* A question is a sentence with a rising inflection at the end. The focal child raising the question must be identifiable at the time the question is asked. When trying to determine whether a sentence is a question, use the responses from the other people to help figure out what the code should be. The other people know the speaker better than we, and they are in the same context in the moment, so they might have a better sense of what the speaker meant.

**Not coded:**

*Unrelated Utterances:* Utterances unrelated to the exhibit will not be coded. (e.g., When do you want to go get lunch? Can you tie my shoe?) Discussions of other exhibits, unless they are relating them to the exhibit, will not be coded.

*Research Awareness comments:* Mentions of the videotaping will be coded as "Research Awareness" and not as a KS or KQ, not as child utterance. See Research Awareness section below.

*Multiple Visitors:* If children other than the focal child speak to the focal child, or anyone else at the exhibit, the non-focal child's utterances will not be coded.

**Coding Scheme Table for Focal Child Questions**

| Code | Type | Description | Examples |
|---|---|---|---|
| KS | Focal Child's Statement | Any statement that has to do with the exhibit. | Answers, explanations, exclamations, grunts, sound effects, singing, etc. |
| KQ | Focal Child's Question | Any question that has to do with the exhibit. | Questions, question fragments |

**Labels for Focal Child Utterances:**

***Content-less:*** Label Kid Statement as content-less if the whole instance lacks semantic content connected to the exhibit and consists solely of meaningless grunts, humming, singing or sound-effects.

***Note:*** Use the note label to mark any code you have a question about, and then use the Text Button to write down the quote, which code you think it might be, and your questions. These notes will be discussed by coders.

ADULTS WITH FOCAL CHILD

## Informative Talk from Adults to focal child

Codes for Informative Talk from Adults to focal child include any information about the exhibit and exhibit use directed to the child. This set of codes does not include questions or any acknowledging, encouraging or discouraging remarks, or statements of opinion about the exhibit (see table below for more detail):

***DN:*** Describe/Name. Descriptions, naming, talk about evidence with no depth to comments.

***DP*:** Directions/Procedure. Direction commenting and procedure commenting; contains information about "how to," but no information about "why."

***ME*:** Meaningful Explanation. Cause and effect; relationships (relating to observed phenomena and more general principles); metaphors/analogies; contains information about "why" something works or does not work.

***UT:*** *Unintelligible Talk*. Can be heard, but not understood (or speaker is unclear). Use this code as a last resort.

***Labels: The two labels for informative talk are Reading Graphics and Notes. Both are discussed in detail following the main code descriptions and examples.***

*Deciding whether an utterance is to be coded as informative talk:* Informative talk provides information about the exhibit to the focal child. The focal adult commenting must be identifiable at the time the informative talk is spoken. When trying to determine whether an utterance is informative talk and when trying to code that utterance, use the context of the interaction to help figure out what the code should be. Informative talk may follow questions or precede responses from the focal child. The other people know the speaker better than we, and they are in the same context in the moment, so they might have a better sense of what the speaker meant.

- Note: Even thought adults often narrate their process as a means of modeling or informing their kid(s) we cannot guess at intent; thus, adult self-talk should NOT be coded as informative talk.
- Adult informative talk should NOT be coded if it is aimed at the self, any one person other than the focal child, or generally at the table such that it is not completely clear that the focal child is one of the targets. Use the context to help determine to whom the adult is speaking. For example, when the adult is responding to a question from someone other than the focal child, it should not be coded. However, when adults are in a back-and-forth conversation with the focal child, even when using I statements, they are coded as speaking with the child—adults are not coded as going in and out of conversation with themselves and the child and back.
- Some cues to look for to decide whether the focal child is a target include: the direction the adult is facing, use of names, whether or not they project their voice, whether they are working together, and whether or not they are responding to or responded to by the focal child.

- If you are unsure of whether the focal child is the target, the default action should be to NOT code the instance.

*Determining whether a fragment can be coded:* In determining whether a fragment can be coded as DN, DP, or ME, listen to the comment preceding and the comment following the fragment to determine whether the context can shed light on the portion of the fragment that is left unsaid during the utterance. Otherwise, do not code.

*Multiple informative utterances:* If a focal adult provides a string of informative utterances without pausing for 2 seconds, code them as a single instance unless there are verbal utterances made by the focal child in between informative utterances (this requires that they are not speaking at exactly the same moment, or overtop of one another completely). If there is a change from informative talk to a Question or vice versa, there should be a break in codes. Qs are not to be part of any informative string. That is, quick changes in speaker, or from informative talk to questions and vice versa, should be broken out into separate codes. If a string contains informative and non-informative talk (other than questions; see Other below), it should be coded as a single string. Each informative talk string should be coded as the highest level of informative talk (see "Determining level" section below). Strings that do not change from adult informative talk to questions are considered an elaboration on the same informative utterance and should be coded as one.

*Determining level of informative talk:* Each informative utterance or string should be coded as its highest level of informative talk; DN is the lowest level of adult information, DP is the medium level of adult information, and ME is the highest level of adult information.

For example: "Look, connect the parts and then it moves," would be coded as Meaningful Explanation (ME), even though "Look" alone or "Look, connect the parts" would be Directions Procedures (DP); however, the inclusion of a causal description, "Look, connect the parts and then it moves," of the sentence includes a Meaningful Explanation which subsumes, or trumps, the DP level for the code.

*Multiple Adults:* All adults with the focal child are coded; however, adult roles are not differentiated. That is, each adult should get their own codes, but the codes will not be separated into enough detail to determine which codes belong to which adult. When two adults are speaking at the same time, both utterances need to be recorded. Because we are interested in counts rather than durations of codes:

- If the adults are speaking at the same time, but the codes are different, code them both at the same time.
- If, however, both adults are speaking at the same time, and their utterances are the same code, be sure to separate the code to account for both instances (because Studiocode combines overlapping instances, two simultaneous codes of the same type need at least a millisecond break in between each code on the same line). This may require the first adult's

utterance to be shortened, or the second adult's utterance to be recorded a few seconds later than its actual occurrence; both instances are fine. You may have to zoom in on your timeline to be able to put a millisecond break in between two codes. Please do not split one adult's utterance in half to place the other adult's simultaneous utterance in the middle of the prior adult's utterance. This breaking up of an instance will cause a single string to be recorded twice. The 2-second multiple utterance rule (see "Multiple utterances" section above) applies to both adults separately. This will be rare.

**Not coded:**

*Other:*

*Unrelated Talk:* Utterances unrelated to the exhibit will not be coded. (e.g., I am going to go to the grocery store after this. I didn't understand that thing over there.)

*Non-Informative Fragments.* A sentence that is not finished or trails off; it cannot be qualified as DN, DP, or ME.

*General Exhibit Talk.* General exhibit talk includes opinions, encouraging or discouraging remarks or statements about the exhibit that do not contain DN, DP, or ME components. If "yes," "no," "uh-huh," "don't know," or "OK" responses to focal child questions that are not seeking DN, DP or ME information; these responses would not be coded.

*Research Awareness comments:* should be coded as "RA," not as adult utterance. See Research Awareness section below.

*Multiple visitors outside the focal child's group:* If visitors outside of the focal child's group inform the focal child, or anyone else at the exhibit, the information will not be coded. When talk is unintelligible, it is often unclear who is speaking (i.e., whether they are in or out of the visitor group); in such cases, do not code the instance. If you are feeling the code belongs to the focal group, code as Unintelligible and use the text box to write your questions about who may be speaking in the text box for that code.

**Labels for Informative Talk from Adults to Focal Child:**

***Reading Graphics:*** When the adults in the focal group read the graphics at the exhibit verbatim (exchanging only articles like "this" or "the"), or they are facing and pointing to the graphics while interpreting them aloud, it will be labeled "Reading Graphics." Depending on what is read, the utterance could be coded as DN, DP, or ME and then also labeled as RG. The Geometry In Motion exhibit labels are located at the back of this appendix—referencing them will aid in determining whether visitors are reading graphics or not.

***Notes:*** Use the "Note" label to mark any code you have a question about, then use the Text Button to write down the quote and your questions; these notes will be discussed by the coders.

## Coding Scheme Table for Focal Adult Informative Talk

| Code | Type | Description | Exemplary | Borderline | Not DN |
|---|---|---|---|---|---|
| DN | Describe/ Name | DESCRIBE, NAME COMPONENTS<br>Reference to the graphics or parts, either by describing or naming, without any elaboration. Statements or descriptions about what can be observed at the exhibit (referencing visual, auditory, or tactile information), with **no** depth or substance about 'how-to,' 'why' or analogous connections.<br>KC[a]: Talk about evidence that can be observed at the exhibit, made reference to visual, auditory, or tactile information that does not establish any causal, analogical or principled connections; DMJ: Description with no depth. | Naming pieces or exhibit components.<br>"There's a shoe" E20607<br>"It's green!"<br>"This Magnet is Strong" | Pointing to a picture on the graphics/label, with some small utterance. While it may be difficult to determine what they are saying, the context of their actions helps determine that they are referencing a component of the exhibit; code as a **DN**.<br><br>"It's a magnet" would be an analogy (or ME) if the exhibit didn't use magnetism to connect the pieces. Given the exhibit design, references to magnets are simply descriptions of an exhibit component, and are coded as **DN**.<br><br>"Here" while holding a piece. They are referencing a component of the exhibit; and if it comes with no further direction (or the surrounding contextual conversation does not imply directions) code as a **DN**. | Naming a picture in the graphics/label, such as windshield wipers in E1, or a horse or train in E2, While these metaphors are suggested in the label, they are still considered a metaphor and thus, an **ME**.<br><br>DN does not include sentences that involve naming of pieces if they can be placed elsewhere, such as "hand me that green one" (**DP**); "the magnets hold them together" (**ME**); or "I like geometry" (**Other= not coded**) |

[a] All "KC" initials were removed from the coding scheme copy provided for one of the coders because she had previously worked in Kevin Crowley's lab.

| | | | | | |
|---|---|---|---|---|---|
| DP | Direction/ Procedure | TRY THIS, DO, HERE'S HOW, ANY DIRECTION, PROCESS<br>Statements providing information about **how to/how not to use** the exhibit or **what to do/what not to do, including quick directives, goal-setting (unless they use a metaphor or analogy), and acknowledgement of what does/doesn't work and the building process**. Includes talk about *own or others* Doing, Using, Trials, or Process.<br>Giving directions on exhibit use that do not delve into meaningful explanations (so do not include any causal, analogical or principled connections). No in-depth information about *why*.<br>KC: giving directions on exhibit use that do not establish any causal, analogical or principled connections; DMJ: Directional commentary; Procedural commenting; no in-depth information about why. | "Oh look, connect the parts" E20607<br>"Oh look, connect the parts like what they did" E20607<br><br>"It almost fits." E10531_vg16<br><br>"Look" E20607<br>"Read the directions." E10531_vg9<br><br>"Let's see"<br>"oh, I see"<br>"here we go"<br>"Stop it" *when talking about using/touching the exhibit*<br><br>"Mine fell apart"<br><br>"Let's try to make something turn."<br>"We're trying to make one"<br>"Build something here" E10524_vg1<br>"Let's make another one"<br>"you make a machine"<br>"Build your own moving sculpture" | "Here" while handing child a piece, where surrounding conversation indicates "use this one", or "put that here"; code as a **DP**.<br><br>Fragments such as:<br>"put that," "you've got to." Or "Maybe we have to…" while not full sentences are enough information to determine that the speaker is providing direction and are thus, coded as **DP**.<br><br>"give it back" or you're hogging all the pieces" counts as **DP** because it is, or is indicative of, a directive about the exhibit pieces.<br><br>"I got it too high." E20531_vg13. This statement is close to an ME because it implies that they mean that it is not working because it is too high (which would be ME), but unless they explicitly state or point out that this is their explanation of why it won't work, we cannot infer their meaning beyond their statement, and thus, we code as a **DP**.<br><br>"I got it" or "you got it" or, these are references to the fact that whatever they've done is working, which is an indicator that what one is doing is correct. Thus, this is a reference to process and is related to what to do; code as **DP**.<br><br>"Watch what mom's doing" is a **DP**, even though it isn't directly about how to use the exhibit, it is giving a direction related to the exhibit and directs the child to observe exhibit use.<br><br>Any response to a kid question about "does this go here" or "is this how it works?" should be coded as **DP**. | "Let's make a sewing machine;" or directing them to grab or make a gear.<br>See E20607_vg13. Both rely on metaphor, which is **ME**.<br><br>"Stop it" *when talking about interpersonal things like hitting or making fun of a brother* (**Other=not coded**) |

| Code | Type | Description | Exemplary | Borderline | Not ME |
|---|---|---|---|---|---|
| ME | Meaningful Explanation | IF, IF/THEN, BECAUSE, IT'S LIKE, AND, CAUSAL MECHANISM<br>Statements providing information about cause and effect; relationships to other exhibits or other/broader phenomena; relationships to more general principles; or using deep or superficial metaphors or analogies (which aid in making connections). May provide information about why something does or does not work.<br>KC: talking about causal connections within the exhibit interface, talking about relations between observed phenomena and more general principles, or about analogies to related phenomena. | "Oh look, connect the parts like what they did and then it moves" E20607<br><br>"It will work better if it's further away" E20531_vg9<br><br>"You're locked up solid because of this one" E20531_vg9<br><br>"This is like the pendulum one."<br>"This is how a piston works."<br><br>References to real world or other conceptual domain. Often using it is or it's like, it reminds me of,<br>Referencing gears or similarity to gears.<br>"…making an arm." (or train, or windshield wipers, or leg, or waving hand…)<br><br>"That's a nice train you built with the magnets and linkages." | Fragments such as:<br>"If …" or "It's just like," or "It won't go all the way, it has to go…" E20607. While not full sentences, these fragments are enough information to determine that the speaker is providing explanation and are thus, coded as **ME**.<br><br>Any response to a "Is this working because of X?" question, or something similar. While it may be difficult to determine the level of what they are saying, if the context clearly suggests that the adult is answering such a question, code as a **ME, even if Yes or No.**<br><br>"The magnets hold them together." While this is not cause-effect about the linkages phenomenon, it is cause-effect about the exhibit interface, and should be coded as **ME.**<br><br>"we're doing geometry" or "it's geometry", or "Geometry in motion" these are coded as **ME** because they are relationships between the exhibit and a broader phenomenological area or more general set of principles (geometry). The last statement would be tagged, **RG**, too.<br><br>Any explanation of magnetism is a discussion of a more general set of principles = **ME** | "we need a magnet" or "put a magnet there" are simply directions not providing enough information to be cause/effect; code as **DP**.<br><br>"That's a nice machine you built with the magnets and linkages"—would be a **DP** b/c it is basically the same as saying build a machine with magnets and linkages. |

## Not Coded

| Code | Type | Description | Exemplary | Borderline |
|---|---|---|---|---|
| Other | The kind of talk we are not coding for includes: general exhibit talk; non-informative fragments; unintelligible talk; and unrelated talk. | Any talk about the exhibit that is not specific enough to be categorized as DN, DP, or ME nor a question. This is talk about the exhibit that cannot be classified, and may seem too meaningful to ignore, but will not be coded for this particular scheme. Often interpersonal, but also includes brief, non-detailed, responses to kid utterances or vague questions, as well as acknowledging, encouraging or discouraging remarks, and statements of opinion.<br><br>Any incomplete or unfinished utterance to focal child where the fragment does not carry enough meaning to be coded as DN, DP, or ME; not enough said to infer type of talk.<br><br>Any talk that is not about the exhibit. | "Grunt"<br>"Oh, man!"<br>"yay"<br>OK<br>Nope<br>Yep<br>Uh huh<br>I don't like this<br>This is hard.<br>I like Geometry<br><br>"good, good, good"<br><br>"I want…"<br><br>"do you want to go get lunch?" | "we can get this" (**Not Coded**)<br><br>"keep trying, you can do it" (**Not Coded**)<br><br>"you'll like this" (**Not Coded**) |

## Adult Questions

**Codes for Questions from Adults to focal child include any question aimed at the focal child about the exhibit. They do not include questions about the E xploratorium or any other unrelated topic (see detailed table below):**

**A*QC*:** Adult Close-Ended Question. This is a question asked allowing only for a yes/no/maybe answer.

**A*QO*:** Adult Open-Ended Question. This is a question asked allowing for a response of some depth further than yes/no/maybe.

**A*QF*:** Adult Question Fragment. A question that is never finished verbally or trails off *and* cannot be qualified as an AQC or AQO. Note that most fragments will be coded as an AQC or AQO.

*Deciding whether a comment is to be coded as a question:* A question is a sentence with a rising inflection at the end. The focal adult raising the question must be identifiable at the time the question is asked (best case is they're on-screen and worst case is they were previously on-screen).

When trying to determine whether a sentence is a question and when trying to code that question, use the responses from the other people to help figure out what the code should be. The other people know the speaker better than we, and they are in the same context in the moment, so they might have a better sense of what the speaker meant.

*Multiple questions:* If a visitor asks a string of questions without pausing for 2 seconds, code them as a single question unless there are verbal responses made by the focal child in between questions, or there is a change in question type (see scheme below), or from a question to informative talk. Strings that do not change are considered an elaboration on the same question and should be coded as one.

**Not coded:**

*Unrelated Questions:* Questions unrelated to the exhibit will not be coded. (e.g., When do you want to go get lunch? Did you return that phone call?)

*Research Awareness Questions:* should be coded as "RA," not as Adult Questions. See Research Awareness section below.

*Multiple Visitors:* If visitors outside of the focal child's group ask questions of the focal child or anyone else at the exhibit, the question will not be coded.

**Labels for Coded Questions from Adults to focal child:**

***Note:*** Use the "Note" label to mark any code you have a question about, then use the Text Button to write down the quote and your questions; these notes will be discussed by the coders.

**Coding Scheme Table for Questions from Adults to focal child**

| Code | Type | Description | Examples | Borderline | Not a Question |
|---|---|---|---|---|---|
| AQC | Close-ended question | ARE, IS, DO<br>Allows for yes/no/maybe answers | "Are these magnets?" E20607<br><br>Do you know what to do here?<br>Do you want some help?<br>Is this interesting to you? | Borderline: E20607_vg13_combo<br><br>Do you know…? | "Do you want to go get lunch?" |
| AQO | Open-ended question | WHAT, WHERE, WHY, HOW<br>Allows for an answer of more depth than yes/no/maybe | "What do you do?" E20607<br><br>"huh" and "what" are vague questions, but still coded as AQO.<br><br>What happens when I do this?<br>How do you make that?<br>Why does it move over there when I turn this? | How does…?<br><br>"would you like to build a hand or a train?" While this question is not completely open-ended, it does allow for more depth than y/n/m questions = AQO. | "What do you want for lunch?" |
| AQF | Fragment of a question | If a focal adult begins to pose a question, but does not complete it (e.g., the person is interrupted), and the fragment does not carry enough meaning to qualify as a AQC or AQO code, or is not interpreted by the focal child in such a way that the meaning of the question can be inferred by the child's answer. The other people know the speaker better than we, and they are in the same context in the moment, so they might have a better sense of what the speaker meant. | Is…? | | |

UNINTELLIGIBLE TALK

**Code for Unintelligible Talk:**

***Unintelligible Talk:*** Utterances made by the focal child or his/her adult(s) that cannot be understood, or the speaker cannot be determined.

**Coding Scheme Table for Unintelligible Talk**

| Code | Type | Description | Exemplary |
|---|---|---|---|
| **UT** | **Unintelligible Talk** | Talk that can be heard but not understood; or understood but speaker cannot be determined. | "nifgetuute"<br>"watch" by someone in the group, but voice cannot be distinguished Code as **UT and mark in text box**. |

These instances will be discussed by the coders.

RESEARCH AWARENESS

**Code for Research Awareness:**

***Research Awareness:*** Utterances made by the focal child or his/her adult(s) that mention the research, videotaping, or microphones will be coded "RA."

*Multiple Research Awareness comments:* If a focal group produces a string of comments regarding the research without pausing for 2 seconds, code them as a single RA instance. *Unlike the other codes, a change in speaker does not require a separate code*, unless there has been a >2 (more than two) second lapse in time. Also, when other codes are simultaneously enacted, or are in the middle of a <2 (less than two) second lapse in the RA talk, the RA code will not be split up but remain continuous, and the other codes will be coded as per usual.

**Research Awareness Coding Scheme Table**

| Code | Type | Description | Examples | Borderline |
|---|---|---|---|---|
| RA | Research Awareness | Any utterances by the focal child and his/her adult(s) pertaining to the research, microphones, videotaping, etc. | "We're being videotaped."<br>"Look" while pointing to mic's.<br>"mommy, what are those black things (mic's)?" | Singing directly following an RA comment (remains an **RA**) |

Coding scheme adapted from two museum studies research projects: APE (Humphrey & Gutwill, 2005) and Powergirl (Crowley & Callanan, 2000 and 2001). Developed by Toni Dancu, Stephanie Bahr, and Dalton Miller-Jones.

[9] This notation was not included in the copies of the scheme provided to coders.

SELECTION FOR VISITOR GROUPS AND ATTRACTION

General Rules:

- You must be looking straight on; do not look at the monitor from any angle.
- Be sure to continually scan both entrances.
- Only look for children, not adults or teens, who fall between the blue and red lines.
    - Clothing, signs of aging, interactions with other visitors and talk can provide clues to whether a visitor is an adult or teen.
- If a child's head grazes under the red line or is within the blue line, they will be "in" (counted) and may qualify as a focal child.
- If a child's head is outside of or falls within the colored portion of the red line or the blue line, they will be "out" (not counted) and cannot qualify as a focal child.
- You must use a child's tallest stance at the point of the red and blue lines even if they are running, jumping, dancing, hair height, or hats. In other words, we cannot assume a child's height by how tall or short they may be without the hat on or if they were standing straight rather than jumping or running. Be sure to see if they stand up straight or walk slowly by any of the height vantage points.
- The bottom line (tape line) on the carpet is to judge where the child should be when looking at their height compared to the red or blue height-lines. It is best to wait until the bottom tapeline is at the child's center of gravity. That is, we are interested in the child's height when their torso or core is hovering over the line – this is most obvious when a child has one foot on one side of the line and the other passing over toward the other side of the line.
- Children who enter from under the cordons (i.e., "break-in") will not usually be eligible for selection. This is because we do not have a good vantage point to estimate their height, unless they pass through the height signs later during their visit. This is also because they are not passing by the signs with a parent, so implied consent is less assured, unless they are with an adult who passes through the cordons. *If they are with an adult who passes through the cordons and they later pass through a height selection point, they can be counted.*
- Children who are already in the area at the start of the taping are not eligible for selection. That is, we are only interested in selecting children who enter and leave the area within the viewable timeframe.
- Children who are clearly part of a camp or school group are not eligible. The best indicators are nametags or matching shirts or uniforms on kids of the same age. However, some families dress their kids in matching shirts, so if you are not sure use their conversation to see if their relationship becomes clear. When in doubt, do not use the group.

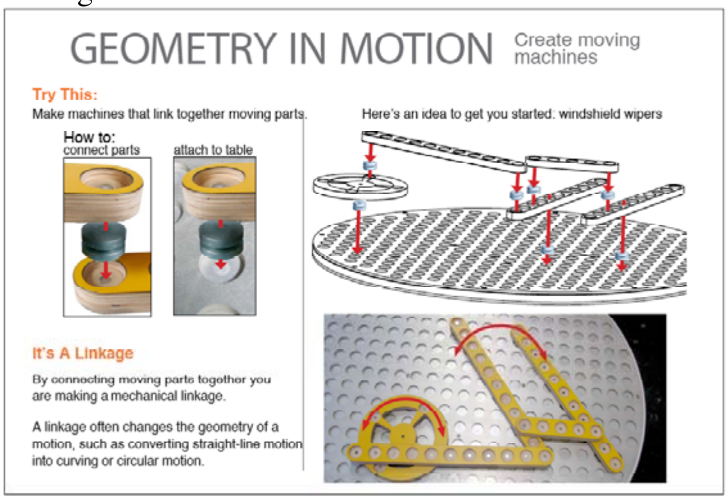
Visitor Group Creation:

- Give precedence to eligible children with adults, versus eligible children alone.
- A visitor group is considered adult and child when both the adult and the child use the exhibit (face feet planted for 3 seconds, or touch), and their use overlaps (occurs simultaneously).
    - It can be difficult to figure out if a particular adult and child who are using an exhibit together are actually visiting together. In such a case, base your decision on: whether they speak with one another, or they provide any clear sign that they are together (hold hands; travel to all 3 exhibits in the area together; speak to each other while at another exhibit…).
- Keep an eye on the discrepancies across groups, if there is a discrepancy greater than five (>5), over-sample the less-represented group until the discrepancy is reduced to 5 or fewer.
- Begin each clip just before the child enters the Sound Abatement area for the first time, and end the clip just after then child leaves the Sound Abatement area for the last time.
- After the child leaves the area, fast forward through the tape for 3 minutes. If they do not return in this time, end the clip after their initial exit. Continue to do this until you reach their final time spent in the area. [Note: If a child returns later, or looks as though you've seen them previously, these incidences should be noted but not included in the clip.]

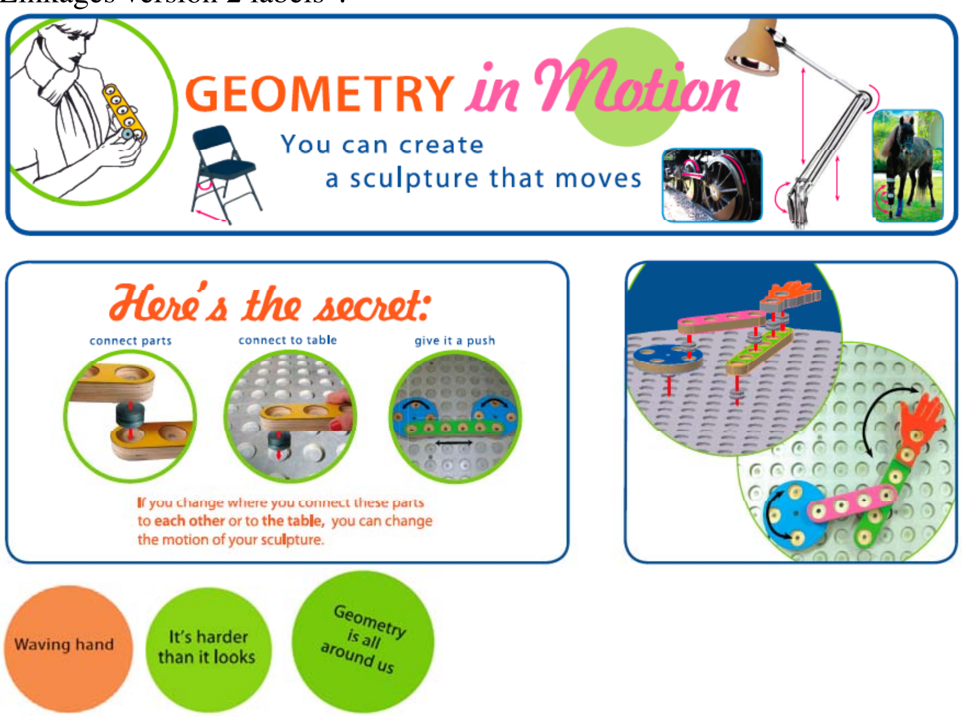
Visitor Group Overlap:

- Visitor groups can be in the SA area at the same time, but they cannot use the exhibit at the same time as anyone in one the focal child's visitor group.
    - If there is overlap, choose the focal child who is part of an adult/child group
    - If the overlapping groups are both adult/child groups, choose the first group to use the exhibit (either adults or children can be initiate use).

Linkages version 1 label:



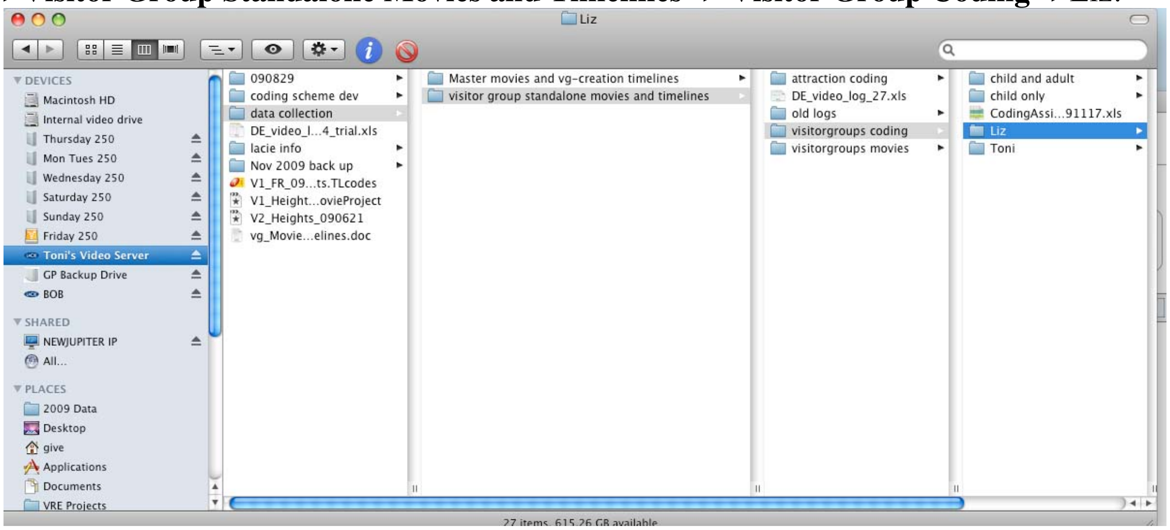

Linkages version 2 labels[9]:



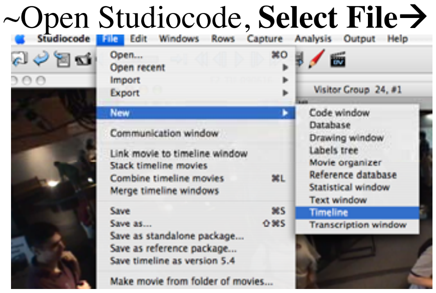

9 The female use drawing was obstructed from view for one of the coders because she had previously worked in the lab of the professor who conducted the Powergirl (Crowley & Callanan, 2000 and 2001) study.

Creating Timelines to Code

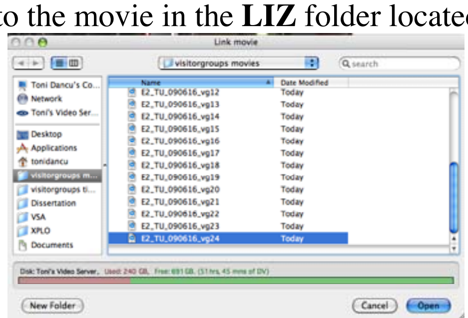
**~Locate your assigned files on** Toni's Video Server **(on the G5) in Data Collection-->Visitor Group Standalone Movies and Timelines--> Visitor Group Coding-->**Liz.[10]




~Open Studiocode, **Select File→ New→ Timeline**

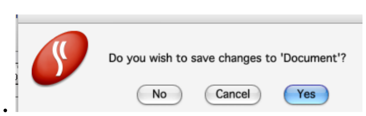



~Link to the movie in the **LIZ** folder located in the **visitor groups movies file**. Press




OPEN.

~Close the timeline, Select **YES** you'd like to save the file.




[10] Note: Each coder had a separate folder by name.

~**Rename** with the same name as the movie file followed by _your initials and save in the **same folder as the assigned movie**.

~Color code in red the file names in the folder of any videos you are not coding.

~Return to YOUR database and enter the reason for not coding in the Reason For Not Coding column.

~At the end of each day, send an email to the other coder indicating how many videos you completed and how many videos remain in the current folder. Do not indicate file names, as many will be co-coded.

~We will meet when each of us has finished the current folder's contents to discuss reliability timelines and unintelligible codes.

~Finalized timelines will be:

1. Color-coded in blue in their respective folders.
2. Copied to **BOB** in the **Final Timelines** folder.
3. Entered into the DE_video_log_mostrecent database as complete.
4. Any videos that are not coded will be tracked by copying our reasons for not coding from our own data bases to the main in the Reason For Not Coding column in the the DE_video_log_mostrecent data base.

Coding Scheme Amendments

1. When talking about photos or videos at the exhibit, the exhibit or parts of the exhibit must be mentioned in order to be counted as exhibit talk. For example: "I want to take a picture of this" would get coded as a KS, but is not informative so would not count for adult talk but as adult Other which is not coded; "Show me how it moves for the video" would get coded as a KS or Adult DP; or "I'll take a picture of your sculpture" would be Adult DN. (December 14, 2009)
2. Metaphorical sound effects (i.e., chuga-chuga-choo-choo, or hello) in reference to movement that animates the linkage effect at the exhibit are coded as superficial metaphors, which are **ME**. Making noises that don't appear to reference real-world noises are not metaphorical or meaningful, and therefore are not coded for adults (January 8, 2010; March 4, 2010)
3. The Reading Graphics label does not apply to interpreting to the pictures (e.g., "see how they put a piece there" or referencing pictures in the labels (e.g., like the train). However, it does apply to adults pointing to the label and remarking "make windshield wipers" or "a waving hand" because those are written aspects of the label and can be read. (February 12, 2010)
4. Breaking apart strings with both question and statement components: If you hear a question linked to a statement, break the question and statement into their respective parts. For example: "Mom! Can I have your wheel?" Can be broken into an exclamation and a question (Mom! = KS and, Can I have your wheel? = KQ), or "You need to put it there, right?" would be broken into two (i.e., right? Is it's own question, while "you need to put it there" is a DP). (February 12, 2010; March 4, 2010)
5. In order to determine whether there are 2 seconds between statements or if they are a single string:
    a. Get the time at the end of the first statement as accurately as possible.
    b. Then move the scrubber ahead 2 seconds:
        i. if you can hear the entire statement from this point forward, code as two instances,
        ii. if you hear anything from the middle of the first word on, code as a single instance. (February 12, 2010)
6. If one of the focal visitors is making a research awareness statement, and they make a codable utterance within their RA talk, code the codable utterance too. For example, if an adult says: " make a machine for the video" we would code it as RA and the portion, " make a machine" as Adult DP. (March 4, 2010)
7. "Huh?" alone is provided as an exemplary open-ended question. However, "huh" following a closed question, such as "you like this, huh?" is coded as closed-ended question because it requires a yes or no answer.
8. Sometimes while using the exhibit, the child may step away to view their or other creations, or walk over to the other side of the exhibit or the parts bin. In these instances, as long as the child is facing the exhibit, they should be coded as engaged (duration is still on), even though their feet are not planted and they are not touching the exhibit at those moments. This exception only applies if they return to feet planted and facing, or touching afterward. (April 7, 2010).

## Appendix E. Visitor Conversations

Below are three excerpts from visitor conversations at the *Geometry In Motion* exhibit. Each example includes a variety of codes that represent how a coded interaction would look.

At the Female-Friendly Featured exhibit, a young girl (FK) and her male caregiver (MC) are building together. In this excerpt, they connect two wheels with a rod, like train wheels.
(visitor group 367)

| Visitor | ***Talk*** **and behavior** | **Code** |
|---|---|---|
| FK | Approaches the second station at the exhibit after having worked for a bit at the first station | |
| FK | *Ooh, look at that!*<br>Spins a wheel | Kid Statement |
| MC | *Yeah, it spins.* | Adult Describe/Name |
| MC | *Now can you make it spin something else?*<br>Grabs another wheel | Closed-ended Question |
| FK | *Yes I can, I need another circle.*<br>Looks to caregiver<br>*I need another circle.* | Kid Statement |
| MC | *OK, here's another circle.*<br>Places his next to hers | Adult Describe/Name |
| MC | *Now you need something to connect them, don't you?* | Closed-ended Question |
| FK | *Yeah, wait a second.*<br>Grabbing a short rod | Kid Statement |
| MC | *I would say something larger.*<br>Grabs a longer rod<br>*I would go for something longer.* | Adult Directions/ Procedure |
| MC | Starts to connect it to her wheels<br>*You have another magnet?* | Closed-ended Question |
| FK | Places a pivot piece on the wheel | |
| MC | Connects the rod to the wheel and turns the wheels round and round using the rod<br>*Look at that.* | Adult Directions/ Procedure |
| FK | *Yeah.* | Kid Statement |

| | | |
|---|---|---|
| MC | *Now if you lower this a little it'll probably work a little smoother*.<br>Lowering one of the wheels on the table, begins turning the wheels with the rod, but it gets blocked by the bell (which is placed on one of the wheels). | Adult Meaningful Explanation |
| FK | *How come it?* | Kid Question |
| FK | Removes the bell<br>*You can't have that there?* | Kid Question |
| MC | Turns the wheels round and round<br>*There you go, that's like a railroad train right there*. | Adult Meaningful Explanation |

| At the Female-Friendly Featured exhibit, a young girl (FK) is working to make the hammer ring the bell. Her male and female adult caregivers go in and out of working with her and watching her build. In this excerpt, the female caregiver (FC) is watching and talking to the girl while she builds.<br>(visitor group 251) | | |
|---|---|---|
| **Visitor** | ***Talk* and activity** | **Code** |
| FK | Working busily, grabbing and connecting parts | |
| FC | *What are you trying to do?*<br>Watching girl | Open-ended Question |
| FK | *Grunts*<br>Still working busily | Kid Statement (contentless) |
| FC | *It's kind of like your Lego toys.* | Adult Meaningful Explanation |
| FC | *Hand?* | Closed-ended Question |
| FK | *I need more hammers.*<br>Looking around | Kid Statement |
| FC | *You need more what?*<br>Still watching | Open-ended Question |
| FK | *I need another hammer.*<br>Looking at other station | Kid Statement |
| FC | *There aren't any other hammers.* | Adult Describe/Name |
| FK | *I need more of these thingy dingys.*<br>Gathering pivot pieces | Kid Statement |
| FK | *Can you take this apart?*<br>Holding a pivot piece attached to a connector up to her caregiver | Kid Question |

| At the Non-Featured exhibit, a young girl (FK) and her male caregiver (MC) are building together. In this excerpt, they connect two separate linkages together. (visitor group 498) | | |
|---|---|---|
| **Visitor** | ***Talk* and behavior** | **Code** |
| FK | *Here, attach this here.*<br>Connects a rod to a wheel | Kid Statement |
| MC | *OK, and make another one?*<br>Attaching a rod to her rod | Closed-ended Question |
| FK | *Here, let's make this over here more.*<br>Moving her rod to a different hole on the wheel<br>*There.* | Kid Statement |
| MC | *Attach it to another wheel?*<br>Grabbing a different wheel | Closed-ended Question |
| FK | *The last wheel, put the last wheel over here.*<br>Grabbing the wheel the MC had and moving it to the other side of the table | Kid Statement |
| MC | *OK.*<br>They both begin working on separate linkages at the table | |
| MC | *Here, I got one to line up.*<br>Attaching the initial rod to a new rod. He tries moving it, but it is locked down. | Adult Directions/ Procedure |
| MC | *So as you turn it, it starts moving things. Right.*<br>He's fiddling with it. Then he is able to turn the wheel and move the rod.<br>*So you want other things to move too. So attach random stuff to it.* | Adult Meaningful Explanation |
| MC | *Something like this?*<br>Grabs more rods to attach to his linkage. | Closed-ended Question |
| FK | *Here.*<br>Attaches pivot piece and rods to his linkage<br>*There.* | Kid Statement |
| MC | *And we can get this and attach this, attach these somehow.*<br>Together they are attaching the two linkages that they made separately<br>*It has to go on the bottom.*<br>Switches two rods<br>*K, now we'll turn this.*<br>Tries to turn wheel but it is locked<br>*We might have to adjust it. This might be a little too big.*<br>Points to their new creation that consists of their separate linkages and begins tinkering | Adult Meaningful Explanation |
| FK | *Oh sec. I'm going to put this [pivot point] here [in the wheel] and see if I can turn it like that.* | Kid Statement |