Responsive Virtual Human Museum Guides

Summative Evaluation

NSF-DRL-ISE # 0813541

April 2012

Prepared for:

Museum of Science, Boston and USC ICT

Prepared by:
Susan Foutz, MA

Jeanine Ancelet, MA

Kara Hershorin, MA

Liz Danter, PhD

Institute *for* Learning Innovation
3168 Braverton St. Suite 280, Edgewater, MD 21037 t: 410-956-5144 f: 410-956-5148 www.ilinet.org
Understanding, fostering, and promoting lifelong learning

This material is based on work supported by the NSF under ISE no. 0813541. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

About the Institute for Learning Innovation:

Established in 1986 as an independent non-governmental not-for-profit learning research and development organization, the Institute for Learning Innovation is dedicated to changing the world of education and learning by understanding, facilitating, advocating and communicating about free-choice learning across the life span. The Institute provides leadership in this area by collaborating with a variety of free-choice learning institutions such as museums, other cultural institutions, public television stations, libraries, community-based organizations such as scouts and the YWCA, scientific societies and humanities councils, as well as schools and universities. These collaborations strive to advance understanding, facilitate and improve the learning potential of these organizations by incorporating free-choice learning principles in their work.

Executive Summary

The University of Southern California's Institute for Creative Technologies (ICT) and the Museum of Science, Boston (MoS) were awarded an Informal Science Education grant from the National Science Foundation (#08133541) for the project, *Responsive Virtual Human Museum Guide.* The goal of the project was to use computer-generated character animation, artificial intelligence, and natural language processing to create interactive characters, or virtual humans, that could engage in face-to-face communication with museum visitors. During the three year project, the MOS and ICT project teams created three exhibits that included or highlighted virtual humans: The Twins, Science Behind, and Coach Mike.

The Institute for Learning Innovation (ILI), a non-profit organization specializing in research and evaluation in informal learning settings, served as the independent, external summative evaluator for the *Responsive Virtual Human Museum Guides* project. The summative evaluation of the project was conducted from May to October 2011. For the summative evaluation, each exhibit was assigned two conditions, allowing for a quasi-experimental design. The evaluation used observations of interactions at the exhibits, interviews with visitors directly after their interaction, and follow-up questionnaires distributed online six week after the museum visit.

The summative evaluation of the *Responsive Virtual Humans Museum Guides* project demonstrated mixed results in incorporating virtual humans into the museum environment. Generally, the Twins and Science Behind (taken together) were successful at achieving the impacts the team identified for the exhibits. The study of Coach Mike, on the other hand, demonstrated that the exhibit was only somewhat successful in achieving the identified impacts. What follows is an overview of the summative evaluation findings, organized by two overarching evaluation questions.

1. **What is the nature of visitors' interactions with the three *Virtual Human* exhibits?**
    a. *Who uses the exhibits?*
    Based on the target visitor who was observed, children were more likely than adults to be the first individual in their group to approach and engage with the exhibits. This finding is in keeping with what has been observed by staff at the museum, with the family's experience strongly shaped by where the children choose to stop.
    b. How do visitors approach and move between the Twins and the Science Behind?
    This study shows that visitors did not frequently move from the Twins to Science Behind, as was originally intended by the project team. The initial idea was that visitors would move naturally from the Twins to the Science Behind to learn more about how the Twins worked; however, this study shows that very few visitors demonstrate that behavior. No summative data was collected on visitor movements from Science Behind to the Twins; in the piloting stage, this type of movement was a rare occurrence and was therefore eliminated as a focus of the summative evaluation.
    c. What are the differences in visitors' interactions with the Twins (types of questions, number of questions, stay time, social interactions) when it is staffed by an MOS interpreter (Blended interaction) verses when it is unstaffed and visitors are interacting with the Twins directly (Direct interaction)?
    There were clear differences between the Blended and Direct interaction approaches tested at the Twins. This is not to say that one type of interaction is better than the other; rather, each approach has unique strengths. This finding supports a flexible

facilitation approach, where staff at Cahners Computer Place can choose to incorporate either Blended or Direct interaction depending on their staffing levels and which outcomes they are looking to support. Specific findings indicate that in the Blended condition, visitors were significantly more likely to: 1) discuss personal and technology/computer-related topics; 2) stay longer at the exhibit (nearly 3 minutes longer); and 3) interact with MOS staff. In the Direct condition, visitors were significantly more likely to interact with other visitors. The facilitation approach (whether Blended or Direct) did not influence the number of verbal utterances visitors addressed to the Twins.

d. *What are the differences in visitors' interactions with Robot Park (stay time, number of programs created, types of programs created, completion of specific tasks, social interactions) when the virtual human guide (Coach Mike) is present and when he is not present?*

   Overall, the presence of Coach Mike at Robot Park did not fully achieve the hypothesized impacts. No significant statistical differences were discovered between the number of programs created, program length, or social interactions when the two conditions were compared. However, some visitor behaviors were influenced positively by the presence of Coach Mike. For example, visitors who engaged with the exhibit with Coach Mike did have significantly longer stay times than those who visited with exhibit without Coach Mike. Coach Mike also supported visitors in completing one of the exhibit challenges (programming the robot to illuminate a sign), as visitors who interacted with Robot Park when Coach Mike is engaged were significantly more likely to complete that goal than are visitors who do not have Coach Mike support. Coach Mike's presence also decreased undesirable behaviors such as 1) using the block tester to move the robot and 2) the writing of long programs instead of editing more, shorter programs. Taking these findings in conjunction with results that show that the current mechanism used to trigger Coach Mike leads to increased usability issues and the highly successful nature of Robot Park without Coach Mike, it appears that the value of Coach Mike is limited to these behavioral improvements, at least in its current iteration.

2. In what ways do interactions with the *Virtual Human* exhibits impact visitors' knowledge and awareness of, engagement and interest in, and attitudes and perceptions towards computer science and technology?

The Twins and Science Behind (taken together) positively impacted visitors across all four impact categories: Engagement and Interest, Attitudes, Awareness, and Knowledge. Overall, visitors found the exhibits engaging, whether the Twins are staffed or unstaffed. Visitors recognized that The Twins are examples of virtual humans and were easily able to identify features of virtual humans. Visitors were curious about the Twins and their capabilities and hold generally positive views about virtual humans in society. As a set of stand-alone exhibits highlighting an advancement in technology and computing, the Twins and Science Behind are successful.

Coach Mike, on the other hand, was not as successful in achieving visitor impacts in the same four areas. This study found evidence for five out of the eleven indicators identified as impact measures; Coach Mike did impact visitors Awareness, Engagement and Interest, and Attitudes for some of the indicators defined for each of these impacts. There was no evidence of a Knowledge impact using the identified indicators. Despite a lack of statistically significant differences between measures such as exhibit ratings, number of programs created, or program length, some visitors who experienced Coach Mike perceived that the intelligent tutor was helpful as a teacher, coach, and guide. There was also evidence that the presence of Coach Mike

might have helped visitors to work with the TERN programming language, significantly reducing behaviors such as using the block tester for the majority of movements, pushing run without the start block, and pushing run without creating a program. There are several ways these results can be interpreted. First, it is important to consider that the Robot Park exhibit existed at MOS before the current project ever took shape. Coach Mike was envisioned as an additional feature at Robot Park that would enhance visitor outcomes at the exhibit. Even before the arrival of Coach Mike, other studies found that Robot Park was a successful, well-designed, intuitive exhibit. Therefore, improving such a successful exhibit is a difficult task. It may be that given the overall high performance of Robot Park, there is little room for overall improvement at the exhibit. Second, usability issues at Robot Park were prevalent when Coach Mike was engaged throughout the duration of this study; had these issues been resolved, visitor impacts relative to Coach Mike might have been stronger. Third and perhaps most important for future work, Coach Mike represents the first known use of an intelligent tutor in an informal education setting. The project team was just beginning their investigation into how this novel addition to a museum might perform; as a result, some of the indicators defined by the team were demonstrated by the evaluation to be unlikely given how visitors interacted with the exhibit and Coach Mike. As the team continues to investigate the integration of virtual humans and intelligent tutors into informal education settings, they will need to continually refine their expectations for what constitutes “success” in these settings.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

The University of Southern California's Institute for Creative Technologies (ICT) and the Museum of Science, Boston (MoS) were awarded an Informal Science Education grant from the National Science Foundation (#08133541) to collaboratively create a life-sized 3D virtual, computer generated character to serve as a museum "educator." The 3-year project, entitled *Responsive Virtual Human Museum Guides,* uses computer-generated character animation, artificial intelligence, and natural language processing to create interactive characters, or virtual humans, that can engage in face-to-face communication with museum visitors. Such characters are not in widespread use, and their appearance in a museum setting is an opportunity to engage visitors in a first-hand experience with cutting-edge science.

The Institute for Learning Innovation (ILI), a non-profit organization specializing in research and evaluation in informal learning settings, served as the independent, external summative evaluator for the *Responsive Virtual Human Museum Guides* project. ILI researchers supported the project teams at ICT and MOS during the first and second years of the project. The summative evaluation of the project was conducted from May to October 2011, the third year of the three-year project. This document reports on the findings of the summative evaluation for the two parts of the project: The Twins, Ada and Grace, and Coach Mike at Robot Park.

Project Background

The *Responsive Virtual Human Museum Guides* was proposed as a project that would create a virtual human to serve as a museum guide in three capacities: 1) as a natural language dialogue-based interactive guide to a MOS gallery, 2) as the basis of a STEM-focused technology exhibit, and 3) as an ongoing research effort that capitalizes on museum visitor participation. The project is designed to enhance a visitor's experience in three ways:

1. The customized and personalized nature of the Virtual Human experience will increase the frequency and depth of visitors' engagement within the exhibition (i.e. Cahners Computer Place);
2. Interacting with the virtual human will encourage social interaction and conversations among visitors;
3. The virtual human and the related AI exhibition and living laboratory will increase visitors' interest in, positive attitudes towards, and comprehension of the STEM science behind virtual humans; i.e., basic computer science, model building, facial and body animation, speech recognition and synthesis, natural language understanding, dialogue management and generation.

The Virtual Human experience is comprised of three exhibits, all of which are the subject of the summative evaluation: 1) the Twins: Ada and Grace, 2) the Science Behind, and 3) Coach Mike at Robot Park. Each of these exhibits is explained below:

The Twins: Ada and Grace

Ada and Grace are virtual humans prominently positioned within Cahners Computer Place, a facilitated gallery at the Museum with a computer and technology focus. Ada and Grace appear as life-sized projections on a large screen (See Image 1). While not photorealistic, their appearance is in-keeping with the advanced graphics the public expects in games for the PC, Xbox or PlayStation. They are women in their twenties, dressed in the red lab coats worn by MOS staff and volunteers. When they speak their voices are heard over speakers and their words are captioned in speech "bubbles" above their heads. Ada and Grace are programmed to understand and respond to questions in a variety of domains, including personal (i.e. "Who are you named after?", "Are you a computer?"), specific exhibits in Cahners, and the STEM content featured in Cahners (i.e. computers, communications, and robots). The Twins receive visitor input as natural language via 1) a table-top microphone with a push-to-talk button that visitors control themselves (i.e. direct interaction) and/or 2) a trained MOS staff member who facilitates the interaction between the visitor and the Twins through the use of a headset microphone as visitors use the table-top microphone.

Image 1: Students meet Ada and Grace at the Museum of Science, Boston. Courtesy Bradley Newman USC ICT.

The Science Behind

This exhibit is directly adjacent to the Twins and demonstrates some of their inter-workings (Image 2). Visitors to the Science Behind can observe the visitors who are currently interacting with the Twins via a video monitor and can see, in real time, the Twins' use of algorithms to determine how to respond to visitor questions. The Twins reference Science Behind to visitors in the course of conversation. For the purposes of the summative evaluation, Science Behind and the Twins were evaluated together, documenting visitor use and impacts across both exhibits.

Image 2: The Science Behind in Cahners Computer Place. Courtesy Catherine Lussenhop, MOS.

Coach Mike at Robot Park

Coach Mike is a virtual human developed to provide scaffolding for visitors at the Robot Park exhibit in Cahners (Images 3 and 4). Robot Park is located opposite the Twins and Science Behind at the far end of Cahners and is not directly related to those exhibits. At Robot Park, visitors use a tangible interface, in this case small blocks each with a specific function, to program a Roomba robot to move or make sounds. Visitors create a sequence of blocks and press the "run" button to send a program to the Roomba. The Roomba can be programmed to complete specific tasks in the exhibit, such as hitting a target to light up a sign. Coach Mike appears on a TV-sized screen on the back wall of Robot Park. He

looks like a cartoon version of a young man wearing a white lab coat. When he speaks, his voice is heard over speakers and his words can be captioned in a speech "bubble" above his head. Coach Mike is an example of an intelligent tutoring system, which are relatively common in computer applications created for use in school-based settings; intelligent tutors give customized support to students as they proceed through an activity or curriculum. However, Coach Mike is the first example of an intelligent tutor within a museum or free-choice learning environment. Coach Mike is programmed to welcome visitors to Robot Park, provide encouragement, make specific suggestions, and through the use of a "Mike" button receive challenges to complete with the robot. He is designed to allow visitors to freely explore Robot Park and his suggestions do not have to be followed by visitors. He receives visitor input through the tangible interface at Robot Park and is aware that a visitor is at the exhibit through the use of a weight/pressure sensitive mat. When a visitor steps on the mat, Coach Mike welcomes the visitor. As Robot Park was an existing exhibit before the *Responsive Virtual Human Museum Guides* project began, visitors are able to program the Roomba with Coach Mike on or off. Coach Mike at Robot Park was evaluated independently of the Twins and Science Behind.

Image 3: Screen shot of Coach Mike. Courtesy USC ICT.

Image 4: Robot Park. Coach Mike appears on the monitor mounted on the back wall of the exhibit. Courtesy Catherine Lussenhop, MOS.

Evaluation Questions and Impacts and Indicators

The summative evaluation focused on answering two primary evaluation questions for the experiences described above. These evaluation questions have been revised from the original NSF ISE grant proposal in response to changes that were made to the experiences during the exhibit development process. The finalized evaluation questions and related sub-questions for each exhibit are detailed below.

1. What is the nature of visitors' interactions with the three *Virtual Human* exhibits?
    a. Who uses the exhibits?

b. How do visitors approach and move between the Twins and the Science Behind?
c. What are the differences in visitors' interactions with the Twins (types of questions, number of questions, stay time, social interactions) when it is staffed by an MOS interpreter verses when it is unstaffed and visitors are interacting with the Twins directly?
d. What are the differences in visitors' interactions with Robot Park (stay time, number of programs created, types of programs created, completion of specific tasks, social interactions) when the virtual human guide (Coach Mike) is present and when he is not present?

2. In what ways do interactions with the *Virtual Human* exhibits impact visitors' knowledge and awareness of, engagement and interest in, and attitudes and perceptions towards computer science and technology?

As required by NSF guidelines, a set of visitor impacts and related indicators were developed and revised by the project team with the support of ILI researchers. These impacts and indicators guided the creation of the evaluation questions, study design and methods, and the development of the instruments. The impacts and indicators are used throughout this report to summarize the findings and are included as Appendix 6.

Methods

Because Coach Mike and Robot Park were independent from the Twins and Science Behind, two separate studies were designed and implemented as part of the summative evaluation. The Twins and the Science Behind were treated as one continuous experience;[1] Coach Mike and Robot Park were a separate experience. To answer the evaluation questions, ILI used a quasi-experimental design to allow for a comparison of visitor usage and impacts between two conditions at each of the two sets of exhibits. The conditions for each exhibit set were as follows:

- The Twins: Ada and Grace and The Science Behind—2 conditions (with no control group):
  1) Interaction with the Twins facilitated via a trained MOS staff member (the Blended condition);
  2) Direct interaction with the Twins, i.e. no staff member stationed at the exhibit (the Direct condition).
- Robot Park—2 conditions (with a control group):
  1) Interaction with Robot Park when Coach Mike is on and providing scaffolding for the visitors' interactions (treatment);
  2) Interaction with Robot Park when Coach Mike is off (control).

The conditions for each exhibit were determined by the MOS and ICT teams, with input from ILI researchers. The goal was to create and test conditions that 1) would be technically possible for the

[1] The decision to treat the Twins and the Science Behind as one experience, or as an exhibit with two components, was based on 1) the inter-related nature of the exhibits and 2) the observed visitor interaction with the exhibits during the formative evaluation and the piloting of the summative instruments.

exhibits, 2) would be logistically possible for the museum staff, and 3) would result in findings that could inform the long-term use of the exhibits in the museum (i.e. the test conditions had to fit within the limits of how staff intended to use the exhibits after the project ended). ILI and MOS staff coordinated to determine when the various conditions would be active and data would be collected. For both exhibits, a schedule was created to ensure that data from each condition would be collected on both weekends and weekdays. For the Twins Blended condition, ILI staff also coordinated with the facilitator of the exhibit, Dan Noren, to create the data collection schedule. Mr. Norn was identified by ITC and MOS as the MOS staff member most familiar with the Twins exhibit, and therefore, most able to facilitate visitor interactions at the exhibit. For Robot Park, MOS staff turned Coach Mike on or off as needed for data collection.

Multiple methods were used to collect data at each exhibit, gathering qualitative and qualitative for each condition. Similar methods were used at both The Twins and Science Behind and Coach Mike at Robot Park, although the types of data collected were tailored to the experience and impacts of each exhibit. Three methods were used:

- Observations: Visitors were observed at each exhibit, allowing for the description of visitors' patterns of usage and experience at each exhibit. The technique of selecting one visitor as the "target" for the observation was used to collect data based on the experience of one individual. A continuous sampling method was used, where once the data collector was in place, the target visitor was the first visitor to stop at the exhibit who appeared to be older than 7 years of age. Data collected at both exhibits included group size and composition, stay time, types of social interaction (between the target visitor and other visitors and between the target visitor and MOS staff/volunteers), and usability issues encountered while using the exhibit. Observations of the Twins also included the number and types of questions asked by the visitor, a categorization of the Twins' responses, the presence of a MOS Staff member as facilitator, and visits to the Science Behind exhibit. See Appendices 1 and 2 for the protocol and observation instrument used at the Twins and Science Behind. Observations at Robot Park included number of programs created, the length of programs created, the completion of the specific tasks that are incorporated into the exhibit, and visitor behavior related to Coach Mike when he was engaged. See Appendices 1 and 3 for the protocol and observation instrument used with Coach Mike and Robot Park.
- Interviews: Interviews were conducted after visitors engaged with either exhibit with the goal of having both an observation and an interview with the same participant. Children under 16 years of age were interviewed only after the data collector obtained permission from an adult family member in the visiting group. Interviews included open-ended questions and rating scale questions for use with all visitors designed to elicit visitor interest, attitudes, awareness, and knowledge of themes related to the visitor impacts. Adult participants only were asked to complete retrospective-pre/post-experience ratings in order to measure change in attitude and awareness as a result of the experience[2]. Demographic and psychographic data were collected from all interviewees (age, gender, zip code, group size and make-up, interest and knowledge of technology and computers). At the end of the interview, visitors were asked to agree to provide contact information for the follow-up questionnaire and were told of the incentive for that questionnaire. A MOS pencil was provided as a thank you gift to the participant and any children

[2] Adult visitors were defined as those 16 years and older. Data collectors determined the age of visitors based on observation to decide which visitors were over 16, and therefore, eligible to complete the retrospective-pre/post ratings. Actual age of the visitor was asked near the end of the interview.

in their group. See Appendices 1 and 2 for the protocol and interview instrument used at the Twins and Science Behind; See Appendices 1 and 3 for the protocol and interview instrument used with Coach Mike and Robot Park.

- Follow-up questionnaire: A web-based questionnaire hosted by Qualtrics was used to gather longitudinal data from visitors who were interviewed and observed at the exhibits. Using contact information collected at the end of the on-site interview, visitors were sent an email invitation to the questionnaire six weeks after their museum visit. The email included a link to the questionnaire and a mention of the incentive. Children under 16 years of age were invited to participate in the online questionnaire through an email invitation sent to their parent's email address; parents were instructed to assist the child in completing the questionnaire. Participants who did not complete the questionnaire within one week of the invitation email received one reminder email. The questionnaire included delayed-post ratings for items included in the on-site interview as well as open-ended questions. At the end of the questionnaire, visitors were asked to provide an email address to be used to deliver the incentive, a $10 gift certificate to Amazon.com. See Appendices 1 and 2 for the protocol and follow-up questionnaire instrument used at the Twins and Science Behind; See Appendices 1 and 3 for the protocol and follow-up questionnaire instrument used with Coach Mike and Robot Park.

All evaluation instruments were developed with the input and feedback of key project staff at both ICT and MOS. The instruments were then piloted by ILI staff, with the necessary revisions shared with the project team. The final instruments and protocol were submitted to the IRB at MOS for approval.

ILI collected the summative evaluation data during the summer of 2011. The Twins and Science Behind data were collected onsite between July 21 and September 11, 2011; online questionnaires were collected between August 20 and October 26, 2011. Coach Mike and Robot Park data were collected onsite between July 1 and August 6, 2011; online questionnaires were collected between August 12 and September 22, 2011. ILI staff members Susan Foutz and Jeanine Ancelet collected a portion of the onsite data. The majority of the onsite data was collected by ILI contractor Emily Shapero, who was trained and supervised by ILI staff. See Table 1 for the sample size for each study.

Table 1: The Sample Size for the Summative Evaluation

Method	Number of Participants	
	The Twins and Science Behind*	Coach Mike and Robot Park**
Observations	225	269
Onsite Interviews	180	223
Follow-up Online Questionnaire	61	75

* The refusal rate for interviews for the Twins and Science Behind was 20%. The response rate for the online questionnaire was 42%.
** The refusal rate for interviews for Coach Mike and Robot Park was 17%. The response rate for the online questionnaire was 40%.

The onsite data for the Twins and Science Behind was entered into Excel and then transferred to SPSS for analysis. The data for Coach Mike and Robot Park were entered into SPSS for analysis. Data from all online questionnaires was downloaded from Qualtrics directly into SPSS. Inductive, emergent-style coding rubrics were created for the analysis of the open-ended data; data from each open-ended

question were reviewed by an ILI researcher and coding categories were created based on themes that emerged from the data. These categories were reviewed by a second researcher and revised based on a common understanding of the categories, data, and intended visitor impacts. A subset of data (20%) was randomly selected based on participant ID numbers. This data was then coded by two researchers using the coding rubrics. A test of inter-rater reliability was performed for each question; questions with less than 70% agreement between the two researchers were reviewed for areas of disagreement. After consultation between the researchers on the areas of disagreement, the rubrics were revised. These revised rubrics were then shared with the project team from ICT and MOS, reviewed, and finalized. The open-ended data were coded and quantified using the finalized rubrics. All data was then analyzed using SPSS; descriptive and inferential (non-parametric) statistical analyses were performed as appropriate. Analysis for the study was undertaken by ILI researchers Susan Foutz, Kara Hershorin, Jeanine Ancelet, and Liz Danter.

Findings: The Twins, Ada and Grace

The findings for the summative evaluation are presented by exhibit, with the Twins and Science Behind presented first, followed by Coach Mike and Robot Park. Within each section, we begin with a description of the sample, followed by findings for each visitor impact area (Engagement and Interest, Attitude, Awareness, and Knowledge) organized by the individual indicators.

Description of Sample

Onsite Observations and Interviews Demographics

A total of 225 observations were completed for the study of the Twins. Of these, 45 were observations with no interviews, and 180 were observations matched with interviews. Participants were evenly split between males and females, with the majority living in Massachusetts (Table 2). Over half of participants (60%) were under 16 years of age; the mean age of participants categorized as "children" was 10.3 years old. Those who were over 16 years old were categorized as "adults" for the purpose of the study; they represented 40% of all participants, with a mean age of 35.3 years.

Demographic and psychographic trends were comparable for both conditions, meaning no statistically significant differences were found between the Direct Interaction and Blended Interaction groups in terms of gender, state/region/country, age, group type, or group size.

Table 2: Onsite Observations and Interview Sample Description

Sample Description	Overall (n=225)	Conditions	
		Direct (n=120)	Blended (n=105)
Gender of Participant			
Male	49%	49%	49%
Female	51%	51%	51%
Live in Massachusetts?	48%	44%	53%
Live in the Northeast?	71%	69%	74%

Live outside of US?	11%	14%	8%
Age of Participant			
Adult (16 and older)	40%	42%	37%
Child (under 16)	60%	58%	63%
Group Type			
Adult only group	13%	13%	13%
Children in group	87%	87%	87%
Mean Group Size	3.4 people	4.0 people	2.9 people

Follow-up Online Questionnaire Demographics

A total of 61 visitors responded to the follow-up questionnaire. As shown in Table 3, about half of the sample was male and half was female. The majority of participants (67%) were under 16 years of age; Adults over the age of 16, represented 33% of all participants. Visitor groups primarily contained children and on average included 5 people. Almost three fourths of the visitors live in the Northeast (67%), with almost half residing in Massachusetts (44%). Overall, the follow-up sample was representative of the onsite sample.

Table 3: Follow-up Online Questionnaire Sample Description

Sample Description	Overall (n=61)	Conditions	
		Direct (n=32)	Blended (n=29)
Gender of Participant			
Male	49%	50%	48%
Female	51%	50%	52%
Live in Massachusetts?	44%	48%	52%
Live in the Northeast?	67%	74%	78%
Live outside of US?	11%	16%	7%
Age of Participant			
Adult (16 and older)	33%	63%	72%
Child (under 16)	67%	37%	28%
Group Type			
Adult only group	13%	19%	17%
Children in group	87%	81%	83%
Mean Group Size	4.8 people	6.2 people	3.3 people

The Twins: Visitor Engagement and Interest

Table 4 shows the Engagement and Interest impacts and indicators for the Twins and Science Behind exhibits. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 4: Twins Engagement & Interest High-Level Results

Impact (Shaded) and its Related Indicators	Level of Evidence Based on the Summative Evaluation
Children (ages 7 -14) and adults will increase their engagement and interest in computer science and technology.	*Achieved: 5 of 5 indicators show evidence of increased engagement and interest.*
Visitors will have **conversations** with other visitors, members of their group, or MOS staff about how the Twins work.	• 87% of participants engaged in social interaction at the Twins or Science Behind. • Conversations about how to operate the exhibits were the most common, followed by conversations about conceptual aspects of the exhibits. • Interactions with staff were more likely in the Blended condition; interactions with other visitors were more likely in the Direct condition.
Visitors will indicate that they had a **positive experience** at the exhibit.	• Participants' median interest rating for the exhibit was 3 out of 4; Blended participants had significantly higher ratings. • Ratings remained high even after six weeks. • Participants indicated that they found interacting with or communicating with the Twins the most interesting aspect of the exhibit.
Visitors will indicate their **interest in learning more about computer science**.	• Participants' median rating for learning about computers by interacting with the Twins was 3 out of 4. • Ratings remained high after six weeks.
Visitors will indicate an **interest to learn more about at least one of the following aspects**: a) how the Twins work, b) other examples of virtual humans, c) other uses for virtual human technology.	• 86% of all participants interviewed demonstrated evidence of the indicator, naming at least one of the identified aspects. • 62% of all participants had additional questions about the Twins after their interaction, with the capabilities of the Twins the most common type of question (35% overall). • 67% of participants mentioned an additional setting where virtual humans could be used.
Visitors will have a **conversation after they leave** the museum about the exhibit experience.	• 54% of respondents to the follow-up questionnaire indicated they did talk to someone after leaving the exhibit, with family members being the most common person to engage in a conversation.

Stay Time at the Twins and Science Behind Exhibits

Participants were timed beginning from when they "stopped" at the Twins exhibit until they left the exhibit; the overall time includes interactions with the Twins, interactions with the Science Behind exhibit, and any interactions with staff or other visitors at either the Twins or Science Behind. Time spent in the exhibit ranged from 19 seconds to just nearly 18 minutes, with a median time of 3 minutes and 7 seconds (Table 5). When looking at stay time between the Blended and Direct groups, those taking part in the Blended interaction were likely to spend significantly more time at the exhibit (Mann-Whitney *U*=3041, N=221, *p*=.000).

Table 5: Stay Time at the Twins and Science Behind Exhibits by the Target Participant

Stay Time	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	221	4 min 14 sec	3 min 7 sec	3 min 36 sec	19 sec	17 min 48 sec
Direct	118	2 min 47 sec	2 min 17 sec	2 min 20 sec	19 sec	15 min 23 sec
Blended	103	5 min 54 sec	4 min 46 sec	4 min 4 sec	40 sec	17 min 48 sec

Description of Usage of and Interactions with the Twins

Participants' and staff interactions with the Twins were observed and recorded by the data collector; data included the number and type of utterances directed to the Twins, who spoke each utterance (the target participant, another visitor, a staff member) and how the Twins responded. Usability issues with the exhibit were also noted.

Visitor and Staff Utterances

The number of utterances recorded for one observation ranged from 1 to 25, with a mean of 8.5 utterances (Table 6). There were significantly more utterances during the Blended interactions than the Direct interactions (Mann-Whitney *U*=4135, N=225, *p*=.000). This difference between the number of overall utterances between Blended and Direct interactions can be attributed to the increase in the number of utterances by staff in the Blended condition; as might be expected there were significantly more staff utterances per observation in the Blended condition (2.8 compared to .2 in the Direct condition; Mann-Whitney *U*=1894, N=225, *p*=.000). There were no differences in the number of utterances by the target participant or other visitors when the two conditions were compared.

Table 6: Number of Utterances Directed to the Twins during an Observation

Utterances	n	Mean	Median	St. Dev.	Min	Max
Overall	225	8.5	7	5.6	1	25
Direct	120	6.8	7	4.4	1	24
Blended	105	10.3	9	6.2	1	25

Utterances were categorized for coding and analysis into four types:

- *Introduction or Greetings*: For introductions and saying hello to the Twins. Includes "Hello", "Hi", "What are your names?"
- *Personal*: For questions/statements addressed to the Twins about themselves that go beyond introductions. This includes "Who are you named after?", "What do you like to do for fun?", "Are you twins?", "What's your favorite color?".
- *Technology or Computer*: For questions/statements addressed to the Twins about technology, computers, and objects or activities in Cahners Computer Place (including robots, artificial intelligence, virtual humans, and cell phones). Includes questions like "Are you a computer?", "What is artificial intelligence?", "Tell me about cell phones.", "Where can I go to learn about robots?".
- *General Museum or Other*: For questions/statements addressed to the Twins about the museum in general (excluding Cahners Computer Place) or other topics not covered by the other categories. Includes "Where are the bathrooms?", "How do I find the dinosaurs?", "What time is the lightning show?", "What is there to see in Boston?", "Will the Red Sox beat the Orioles?"

When looking at the utterances by these categories a few trends emerged (Table 7). Utterances of a personal nature were the most common type of utterance directed to the Twins (with a mean of 4.0 personal utterances per observation) followed by utterances related to technology or computers (mean of 2.2). There were significant differences between the conditions, with the Blended condition having a greater number of both personal and technology/computer-related utterances.

Table 7: Utterances in each Category by Condition

Challenge	Mean			*U* value	*p* value
	Overall (n=225)	Direct (n=120)	Blended (n=105)		
Introduction/Greeting	1.5	1.4	1.7	5806.5	.287
Personal	4.0	3.1	5.0	4216.5	.000**
Technology/Computers	2.2	1.6	2.9	4621.0	.000**
General Museum/Other	0.7	0.8	0.7	6280.0	.962

Note. The Mann-Whitney U test was used to test for statistical significance.
* Significant at the p<.05 level. **Significant at the p<.01 level.

Data collectors also noted 1) whether any of the utterances of the target participant or other visitors was one of the sample questions listed on the label next to the microphone and 2) if they repeated an utterance exactly. More than half of all visitors (56%) asked a question that was on the label. As may be expected, visitors were more likely to ask questions listed on the label in the Direct condition (χ^2= 7.656, N=225, p=0.006); the mediation of the staff member in the Blended interaction lessened visitors' reliance on the label. More than half of all visitors also repeated an utterance exactly (54%). Repeating a question exactly was usually a reaction to the response of the Twins; visitors would repeat a question if they did not receive an appropriate answer (as defined below). Visitors in the Blended condition were more likely to repeat a question (χ^2= 4.077, N=225, p=0.043); this is related to the mediation of the staff member. When a visitor did not get an appropriate response to their question, the staff member would

encourage the visitor to ask it again, often giving advice on how to use the exhibit (i.e. speak slower, come closer to the microphone).

The Twins' Responses

The responses given by the Twins to each utterance were also recorded and categorized into three types:

- *Appropriate:* The Twins response to the utterance was a "reasonable" answer to the question or statement posed.
- *Inappropriate:* The Twins response to the utterance was an "unreasonable" answer to the question or statement posed. Unreasonable means off-topic or unrelated to what was asked.
- *Don't Know*: The Twins response indicated that they did not know the answer to what was asked. The cause of this response could be because it was outside of their knowledge domain or because they did not understand the questions. The Twins had a variety of ways to indicate that they did not know, including a straightforward response and jokes.

The Twins were also programmed to give suggestions of other topics to ask them about or related areas or exhibits in Cahners as part of their response to visitors' questions; these suggestions were also recorded.

Overall, the majority of responses given by the Twins were appropriate to the questions they were asked (mean of 4.8 appropriate responses per observation). This was followed by "don't know" responses (mean of 3.0) and inappropriate responses (mean of 0.6). Contrary to what might be expected, the mediation of a staff member in the Blended interactions did not reduce the overall number of "don't know" or inappropriate responses; these were the same in both conditions. There were significantly more appropriate responses in the Blended condition as compared to the Direct condition (a mean of 6.5 compared to 3.3; Mann-Whitney U=3325.5, N=225, p=.000); this increase is attributed to the overall greater number of utterances directed towards the Twins during Blended interactions.

The overall number of suggestions the Twins gave during each observation was relatively low with a total of 128 suggestions given over 225 observations, a mean of 0.6 suggestions per observation. More than a third of all suggestions (37%) directed the visitor to the Science Behind exhibit and 23% directed visitors to 20 Questions. The Twins were more likely to make a suggestion in the Direct condition (Mann-Whitney U=5388, N=225, p=.034), but there were no differences between conditions in their likelihood of suggesting the visitor go to Science Behind.

Usability Issues at the Exhibit

General usability issues at the Twins exhibit were also recorded. Overall, 45% of all groups observed experienced a usability issue of some type. The most common usability issues included misuse of the microphone button (occurring during 29% of all observations) and speech issues (24% of all observations) (See Appendix 1 for a definition of the usability issues). It should be noted that the noting of a usability issue was based on the data collector's discretion and understanding of the exhibit. In the Direct condition, visitors likely would not realize they were having a usability issue; instead it would be perceived as a misunderstanding on the Twins' part. In the Blended condition, the staff member could call the visitors' attention to a usability issue in an attempt to have them correct it. When the overall incidence of usability issues in the two conditions was compared, there were no significant differences.

There were significantly more occurrences of button misuse in the Blended interaction as compared to the Direct interaction (χ^2= 3.863, N=225, p=0.049); however, this difference may be a result of the staff member calling the attention of both the visitor and the data collector to the misuse. There were no differences in the occurrence of usability issues between adults and children.

Description of Usage of and Behavior at Science Behind

During the observations, data collectors noted whether the target participant of the observation visited Science Behind. The observational protocol included only those visitors who started at the Twins and then went to Science Behind; visitors who started at Science Behind were not included in the study. Therefore, the study captured the relationship between seeing the Twins first and then visiting Science Behind, and not the reverse.

Of the 225 target participants observed, only 39 (17%) visited Science Behind after interacting with the Twins. Significantly more participants in the Direct interaction condition visited Science Behind (23% compared to 11% of Blended visitors; χ^2= 6.460, N=225, p=0.011). In the Blended condition the staff member supplied information about how the Twins work, so it makes sense that without a staff member present visitors would seek out more information from Science Behind. Visitors to Science Behind who had just interacted with the Twins tended to look at the screen showing the real time input/output from the conversations the Twins. Many of these visitors looked for the questions they had asked the Twins to see how it was interpreted. Some visitors would visit Science Behind while another member of their group was interacting with the twins. There were no significant differences between how often adults and children visited Science Behind.

As described above, the Twins suggested that visitors go to the Science Behind exhibit; a more than third of all suggestions made by the Twins (37%, n=47) were related to Science Behind. Based on condition, there were no differences in how often the Twins suggested a visit to Science Behind. However, there was a relationship between the suggestion of going to Science Behind and actually going; participants in interactions where the twins suggested going to Science Behind were significantly more likely to visit it than those who had not heard this suggestion (32% compared to 14%; χ^2= 8.816, N=225, p=0.003).

Visitors will have conversations with other visitors, members of their group, or MOS staff about how the Twins work (Engagement and Awareness Indicator)

The measure of social interaction was based on the interactions of the target participant being observed with any MOS staff (or volunteers) and other visitors to the museum. Only verbal interactions were noted by the data collector. Both interactions initiated by the participant and directed to the participant were included. Social interactions were observed at both the Twins and Science Behind. At both locations, social interactions were categorized by type (described below) for analysis.

Social Interactions Overall

When considering social interactions at both the Twins and the Science behind combined, the majority of participants (87%) interacted with either a staff member or another visitor. More than half of all participants (51%) interacted with a staff member and two-thirds (66%) interacted with another visitor. Social interaction with either a staff member or another visitor was linked to an increased stay time for the target participant (Table 8). Participants who interacted with a staff member stayed on average 3

minutes and 12 second longer than those not interacting with a staff member. Participants who interacted with another visitor stayed on average 1 minute 16 second longer than those not interacting with another visitor.

Table 8: Stay Time by Social Interaction

Social Interaction	Stay Time: Mean Rank		Sample Size		*U* value	*p* value
	Interaction observed	Interaction not observed	Interaction observed	Interaction not observed		
Interaction with Staff? (N=216)	136.14	78.73	112	104	2728.0	.000**
Interaction with Another Visitor? (n=220)	121.17	90.28	114	76	3935.0	.001**

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. **Significant at the p<.01 level.

The type of social interaction observed was also linked to the condition type. As expected, interactions with staff were much more likely to occur under the Blended condition (χ^2= 132.840, N=220, *p*=0.000); 93% of those in the Blended condition spoke with a staff member compared to only 15% of participants in the Direct condition. Interaction with another visitor was more likely to occur in the Direct condition (χ^2= 8.359, N=224, *p*=0.004); 74% of those in the Direct condition spoke to another visitor compared to 56% of those in the Blended condition. There were no significant differences between adults and children were looking at overall interaction with a staff member or with another visitor; adults and children interacted with others in similar amounts.

Social Interactions at the Twins

Of the 223 people whose social interactions at the Twins were observed, 190 (85%) had a social interaction with either another visitor or a staff member at the exhibit. Overall, interactions with other visitors were more common than interactions with staff (Table 8). Mirroring the results seen in the combined data from the Twins and Science Behind (reported above), interactions with staff at the Twins was much more likely to occur under the Blended condition than in the Direct condition (χ^2= 135.659, N=223, *p*=0.000). Similarly, interactions with other visitors at the Twins were more likely to occur under the Direct condition (χ^2= 4.862, N=224, *p*=0.027).

Social interactions also were categorized by the data collector as they occurred into the following types:

- *Conceptual*: Conceptual talk about the purpose of the exhibit or the concepts the exhibit is trying to teach. The talk may or may not be sophisticated or not; for example “Here you get to talk to these girls” is less sophisticated than “Here you can talk to the virtual humans Ada and Grace” but both are examples of conceptual speech. Also includes general questions such as “What is this about?” or “What do you do here?”
- *Operational*: Talk about how to operate the exhibit or what specifically to ask the Twins. Utterances may be *instructions* (“Push the button”, “Sit on the stool”, “Lean closer to the microphone”, “Ask them about cell phones”), *suggestions* (“You could sit on the stool”, “Try pushing the button”, “Ask them something you would ask a friend”) or a *question* (“What do

you think this button does?", "What will happen when I talk to them?", "What do they know about?"). Also includes statements such as "I don't know what to ask" and "Tell me what to ask."

- *Positive Affect*: Any positive talk or utterance. Includes "Cool!", "They understood!", and laughing.
- *Negative Affect*: Any negative talk or utterance. Includes "It's not working", "They are dumb", "Boo!"
- *Other Social Interaction*: Any verbal interaction by the participant with person that does not clearly fit into one of the categories above should be explained in this category. This includes instances where the visitors cannot be clearly heard but can be seen to be talking and instances where the visitors are speaking a language that the data collector does not know.

Of the interactions recorded, interactions with staff tended to focus on operational and conceptual issues, 47% and 32% respectively (Table 9). Interactions with other visitors tended to be either an expression of positive affect (41%) or operational in nature (38%).

Table 9: Social Interactions at the Twins by Type

Type of Social Interaction	Interaction with Staff (n=223)		Interaction with Another Visitor (n=224)	
	n	Percent*	n	Percent*
Overall	114	52%	142	63%
Conceptual	71	32%	35	16%
Operational	105	47%	86	38%
Positive Affect	55	25%	92	41%
Negative Affect	7	3%	22	10%
Other Social Interaction	4	2%	2	1%

* Multiple types of participation could occur. Percents total more than 100%

Analysis by type of social interaction yielded the following significant results:

- Participant interactions with staff:
  - When comparing the interactions of **adult and child participants** with staff, children were more likely than adults to talk about **operational issues** with staff (χ^2= 4.681, N=185, p=0.030).
  - Participants who engaged in the following types of interactions with staff were likely **to spend more time** at the Twins: **conceptual** (Mann-Whitney U=2925, N=220, p=.000), **operational** (Mann-Whitney U=3019.5, N=220, p=.000), and **positive affect** (Mann-Whitney U=1657.5, N=220, p=.000).
- Participant interactions with other visitors:
  - When comparing the interactions of **adult and child participants** with other visitors, children were more likely than adults to talk about **operational issues** with another visitor (χ^2= 5.095, N=185, p=0.024).
  - Participants who engaged in **positive affective** speech with another visitor were likely **to spend more time** at the Twins (Mann-Whitney U=3682.5, N=220, p=.000).

Social Interactions at Science Behind

Of the 39 people who were observed using Science Behind, 19 (49%) had a social interaction at the exhibit. Interactions with other visitors were more common than interactions with staff (Table 10). Of the interactions recorded, conceptual interactions were the most common type, followed by interactions about operating the Twins. The overall number of social interactions at Science Behind was too small to merit the use of inferential statistics.

Table 10: Social Interactions at Science Behind by Type

Type of Social Interaction	Interaction with Staff (n=39)		Interaction with Another Visitor (n=39)	
	n	Percent*	n	Percent*
Overall	6	15%	13	33%
Conceptual	5	13%	10	26%
Operational	4	10%	8	21%
Positive Affect	2	5%	4	10%
Negative Affect	2	5%	1	3%
Other Social Interaction	0	--	0	--

* Multiple types of participation could occur. Percents total more than 100%

Visitors will indicate that they had a positive experience at the exhibit (Engagement and Awareness Indicator)

In order to determine whether visitors had a positive experience at the exhibit, visitors were asked a series of open-ended and close-ended questions. In order to gather emergent, top-of-mind responses, visitors were asked the open-ended question: "What did you find most interesting about the exhibit?" As a way to triangulate, visitors were then asked to rate a series of statements related to their experience with the exhibit overall and with specific functions of the exhibit, as well as with their desire to interact with the exhibit further. In each section below, the overall frequencies are reported first, followed by a discussion of any significant differences among the sample according to the following variables: 1) type of interaction; 2) age; 3) time; 4) knowledge and interest scales; 5) visitation to Science Behind; and 6) usability issues.

Open-ended Expression of Interest in the Exhibit

As an opening question to gauge visitors' initial reactions to their experience at the exhibit, visitors were asked what they found most interesting about the exhibit. Responses to this question were coded and quantified by ILI researchers into six broad categories: 1) Interacting and communicating with the Twins; 2) Thought-provoking nature of the exhibit; 3) Quality of the design and uniqueness; 4) Watching others interact with the exhibit; 5) Learning something new about science, technology, and/or computer content; and 6) nothing was interesting. A seventh category was created to capture responses that were either un-codeable or fell into some other category.

Just over two-thirds of visitors (67%) indicated that the most interesting aspect of the exhibit was having the opportunity to interact and/or communicate with the Twins (Table 11). Notably less

visitors talked about other interesting elements, including the quality of the design (19%), learning something new (7%), the thought-provoking nature (5%), and/or watching others use it (2%). Only a few people (5%) said that nothing was interesting; most of these people expressed frustration that they could not get the Twins to work and were disappointed in the Twins' ability to understand and respond to their questions.

Table 11: Responses to "*What did you find most interesting about the exhibit?"* (n=180)

Responses	n	Percent*
Interacting/communicating with the Twins	120	67%
The interactivity (i.e. asking a question and/or getting a response	68	30%
Their ability to understand, their accuracy, and/or limits	46	20%
Their personality and/or sense of humor	8	4%
Quality of the design / Uniqueness	34	19%
Learn something new about science, technology, and/or computer content	13	7%
Thought-provoking	9	5%
Watching others use it	3	2%
Nothing	9	5%
Other / un-codeable response	11	6%

*Multiple responses allowed. Percentages total more than 100%.

Interacting/communicating with the Twins: The most commonly mentioned aspect of the exhibit focused on interacting or communicating with the Twins (67% of respondents). This category was divided into three sub-codes to better understand what aspect of the interaction visitors found interesting. The largest subset of visitors focused on the interactivity of the exhibit, which was focused on asking the Twins questions and receiving a response (30% overall). Many of these visitors enjoyed the personal nature of the exchange, indicating that they enjoyed talking to Ada and Grace. Another sub-set of visitors focused their response on the Twins' ability to understand, their accuracy, and the limits of that understanding or accuracy (20% overall). These visitors were interested in the Twins' ability to understand different people and to respond to a variety of questions. A few visitors (n=4) specifically mentioned the humor embedded in the Twins' responses or other aspects of their personalities (n=4). Some examples of visitors' comments include:

I like talking to them.
They were able to communicate like normal humans.
They could understand different ways of saying things.
They're programmed to respond to a lot of things which is cool.
Trying to figure out what questions they could and understand and which they couldn't.
Their responses are funny.
They are witty! You find out what they like.

Quality of the design/ Uniqueness: Around a fifth of visitors commented on the quality of the design and/or the uniqueness of the exhibit as being the most interesting aspect. Many of these visitors commented on the Science Behind, saying that it was an important element of their experience,

while others commented on the sophistication of the graphics and life-like quality of the Twins. Other people focused on the uniqueness of the exhibit, saying things like "I've never seen anything like it before."

Learning something new: The findings for the overall summative evaluation demonstrate that diverse and wide-ranging learning was taking place at the exhibits. However, learning as a top-of-mind "interesting" aspect of the exhibits was mentioned infrequently. A few people talked specifically about what they learned at the exhibit as the most interesting aspect, usually naming specific facts that they learned. A majority of these visitors specifically mentioned learning about how the Twins were named after Ada Lovelace and Grace Hopper. Other visitors said that they enjoyed learning something new about computers or virtual humans in general.

Thought-provoking: Somewhat related, a few people said the most interesting aspect of the exhibit was the questions it raised for them, or in other words, the way it provoked them to think of something new and stimulated curiosity. A few people said the exhibit made them think about the possible uses of virtual humans and the costs and benefits associated with the technology. Others pondered the technology required to build Ada and Grace, saying things like "How much programming did it take," and "just trying to figure out how it worked."

When each category of response was compared by type of interaction (Blended or Direct), significant differences were found.

- Visitors who interacted with the Twins under the Direct condition were significantly more likely to say that **nothing was interesting** about the exhibit (χ^2= 7.579, N=180, p=0.006).
- When comparing between response categories and age group (adults and children), differences were also found among visitors who said nothing was interesting; children were significantly more likely to say that **nothing was interesting** about the exhibit than were adults (χ^2= 4.789, N=176, p=0.029).

Significant differences were also discovered when each category of response was compared by time spent at the exhibit.

- Visitors who spent shorter amounts of time at the exhibit, were significantly more likely to say **nothing was interesting** about the exhibit (Mann-Whitney U=451.5, N=180, p=0.037).
- Usability also played a factor in determining how visitors responded to this question. Visitors who experienced usability issues were significantly less likely to say that what they found most interesting was **interacting with and/or communicating with the Twins** (χ^2= 4.949, N=180, p=0.026).

No significant differences were found when comparing response categories with visitors' knowledge and interest, or their visitation to the Science Behind.

Rating of their Experience at the Exhibit

A quantitative rating scale question was also used to determine if participants had a positive experience at the exhibit. Participants were asked to rate their interest in "Interacting with the exhibit" on a four point scale, where 1 was "boring" and 4 was "exciting." As shown in Table 12, the overall rating for all participants was a median of 3, or "pretty good" on the scale. There was a statistically significant difference in the ratings of the Blended and Direct participants, with Blended participants rating their interest higher (Mann-Whitney U=2998, N=174, p=0.011). Table 13 breaks down the frequencies for each rating.

Table 12: Interacting with the Twins and Science Behind Rating

Interacting with the exhibit was…	N	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	174	3.07	3	.645	1	4
Blended	76	3.21	3	.618	1	4
Direct	98	2.97	3	.649	1	4

Table 13: Interacting with the Twins and Science Behind Frequencies

Rating	Overall (n=174)	Conditions	
		Blended (n=76)	Direct (n=98)
Boring	1%	1%	1%
Just Okay	14%	7%	19%
Pretty Good	62%	62%	61%
Exciting	24%	30%	18%

Other significant findings for the rating of "Interacting with the exhibit" include:

- Children rated their interactions with the exhibit **higher** than did adults (Mann-Whitney *U*=2743, N=174, *p*=0.008).
- There was a positive correlation between a high exhibit ratings and **spending more time** at the exhibits (Spearman's Rho=.214, N=173, *p*=.005).
- Those participants who had a **social interaction with another visitor** tended to rate their interaction with the exhibit higher (Mann-Whitney *U*=2631.5, N=174, *p*=0.049).
- Those participants who had a spoke with **staff about an operational aspect** of the exhibit tended to rate their interaction with the exhibit higher (Mann-Whitney *U*=3081, N=173, *p*=0.022).

Participants were asked this same question six weeks later in the follow-up online questionnaire; researchers compared ratings from the interview and follow-up questionnaire. Ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Visitors will indicate their interest in learning more about computer science (Engagement and Interest Indicator)

A quantitative rating scale question was also used to determine if participants had a positive experience at the exhibit. Participants were asked to rate their interest in "Learning more about computers by interacting with the Twins" on a four point scale, where 1 was "boring" and 4 was "exciting." As shown in Table 14, the overall rating for all participants was a median of 3, or "pretty good" on the scale. There was no statistically significant difference in the ratings of the Blended and Direct participants. Table 15 breaks down the frequencies for each rating.

Table 14: Learning about Computers by Interacting with the Twins Rating

Learning more about computers by interacting with the Twins would be…	N	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	174	2.94	3	.803	1	4
Blended	76	2.93	3	.822	1	4
Direct	98	2.95	3	.791	1	4

Table 15: Learning about Computers by Interacting with the Twins Frequencies

Rating	Overall (n=174)	Conditions	
		Blended (n=76)	Direct (n=98)
Boring	2%	1%	3%
Just Okay	28%	33%	25%
Pretty Good	43%	37%	47%
Exciting	27%	29%	26%

Other significant findings for the rating of "Learning more about computers by interacting with the Twins" include:

- There was a positive correlation between a high interest ratings and **having a high self-reported knowledge of technology** (Spearman's Rho=.192, N=174, *p*=.011).
- There was a positive correlation between a high interest ratings and **having a high self-reported interest in technology** (Spearman's Rho=.228, N=174, *p*=.002).
- There was a positive correlation between a high interest ratings and **having a high self-reported interest in computers** (Spearman's Rho=.207, N=174, *p*=.006).

Participants were asked this same question six weeks later in the follow-up online questionnaire; researchers compared ratings from the interview and follow-up questionnaire. Ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Visitors will indicate an interest to learn more about at least one of the following aspects: a) how the Twins work, b) other examples of virtual humans, c) other uses for virtual human technology (Engagement and Interest Indicator)

Participants in the interviews were asked what additional questions they had about Ada and Grace and in what other settings virtual humans might be useful. The data from these open-ended questions were used to create three quantitative variables for additional analysis. These qualitative and quantitative results are reported below.

Questions about the Twins after the Interaction

Participants were asked "What questions do you have about Ada and Grace or what else would you like to know about them?" Participant responses to this question were categorized into six categories: 1) Initial idea and implementation, 2) capabilities of the Twins, 3) future outlook for the Twins, 4) personification of the Twins, 5) other questions or a request for how to find out more, and 6) no further questions.

About a quarter (26%) of all participants did not have any questions (Table 16) after interacting with the exhibit. There were no differences between condition (Blended or Direct) or by age (adult or child) with regards to having no further questions about the Twins. Of those participants who did have questions, the most common questions were related to the capabilities of the Twins (35%), followed by questions related to the initial ideas for creating the Twins and how they were created (29%). Each of the categories is explained in more detail below.

Table 16: Responses to *"What questions do you have about Ada and Grace or what else would you like to know about them?"* (n=171)

Responses	n	Percent*
No questions	44	26%
Capabilities of the Twins	59	35%
Initial Idea & Implementation	49	29%
Personification of the Twins	14	8%
Other or request for more information	13	8%
Future Outlook for the Twins	4	2%

* Multiple responses allowed. Percentages total more than 100%.

Capabilities of the Twins: The most common type of question about the Twins was related to what they can do and how they work. Visitors were interested in the boundaries of the Twins' knowledge and abilities. Representative examples of this type of question included:

What kinds of questions can they answer?
How many words would they know and how many sentences can they understand?
Do they know everything that everybody knows [and] do they know who used to be the president?

Initial idea and implementation: Many visitors also wondered about who thought of or created the Twins, how were they made, how long it took to make them, and where they were created. Others wondered how long they had been at the museum or how much they cost to make. Others wondered if the Twins were based on real people.

Personification of the Twins: A few visitors wanted to know more about aspects of the Twins that may only be true of real people. These visitors were personifying the Twins. For example, one child wondered if the Twins were identical.

Other or request for more information: A small number of visitors specifically wanted to know where they could get more information about the Twins or concepts such as AI. Adults were significantly

more likely to ask where they could go for more information than children (χ^2= 10.433, N=171, p=0.001).

Future outlook for the Twins: A few visitors wanted to know what would happen to Twins, such as how much longer they would be at the museum.

When each category of response was compared by type of interaction (Blended or Direct), no significant differences were found. This is an interesting finding, in that participants in the Blended condition had the opportunity to ask the staff member facilitating the interaction questions, where as visitors in the Direct did not. Despite this opportunity, participants who had ready access to a staff member still had questions about the Twins. It could be that the interview process itself helped all visitors to generate more questions about the Twins.

Other Settings Where Virtual Human Might Be Useful

Participants in the interview were asked if they could think of other settings where virtual humans could be useful. Responses were categorized into four areas of usage (hospitality, instructional/educational, dangerous, and medical) and additional three categories (other uses not fitting into the above categories, no other possible uses, and "I don't know"). As seen in Table 17, the most common response was in the hospitality industry (36%), followed by instructional or educational settings (27%). A small proportion of those asked (10%) indicated that there were no other possible uses for virtual humans; the response was generally related to the visitor feeling the Twins were not very sophisticated and therefore had limited applicability.

Table 17: Responses to *"Can you think of other setting where virtual humans might be useful?"* (n=170)

Responses	n	Percent*
No additional uses possible	17	10%
Hospitality (Travel, Entertainment, Food Service)	61	36%
Instructional / Educational settings	45	27%
Dangerous or Hostile environments	11	7%
Medical Industry	11	7%
Other Uses	28	17%
"I don't know"	15	9%

* Multiple responses allowed. Percentages total more than 100%.

Additional analysis revealed the following areas of significant differences were found:

- Blended participants were more likely to mention an **instructional/educational usage** (χ^2= 6.754, N=170, p=0.009). Mentioning instructional or educational uses was tied to having **spent more time** at the exhibit (Mann-Whitney U=1997, N=169, p=.005).
- Direct interaction participants were more likely to mention **hospitality-based uses** (χ^2= 4.466, N=170, p=0.035), as were adults (χ^2= 15.234, N=170, p=0.000).

- Children were more likely that adults to feel that there were **no other uses for virtual humans** χ^2= 4.976, N=170, *p*=0.026).

Quantitative Variables as Indicator of Interest

Using the responses to these open-ended questions, three variables were created to determine if evidence of the following indicators was found: Visitors will indicate an Interest in learning more about a) how the Twins work, b) other examples of virtual humans, or c) other uses for virtual human technology.

- Had a question relative to the indicators. A new variable was created by combining the responses of participants who had a question about how the Twins work, other examples of virtual humans, or other uses. Using this variable, 62% of all participants who were interviewed had a question that was applicable to the indicator. There were no significant differences based on condition (Blended or Direct) or age (adult or child participant).
- Was able to identify an additional setting for using virtual humans. A variable was created by combining the responses of all those who mentioned another setting in which to use virtual humans. An analysis of this variable indicated that 67% of those interviewed were able to mention another setting. Adults were significantly more likely that children to have mentioned another possible setting χ^2= 14.998, N=170, *p*=0.000). There were no significant differences based on condition (Blended or Direct).
- Overall evidence of the indicator: A third variable was created by combining the two above variables. Analysis of this variable reveals that 86% of all participants interviewed demonstrated evidence of the indicator overall. There were no significant differences based on condition (Blended or Direct) or age (adult or child participant) for this variable.

Visitors will have a conversation after they leave the museum about the exhibit experience (Engagement and Awareness Indicator)

Follow-up questionnaire participants were asked the closed-ended question: "After you left the museum, did you talk to anyone about Ada and Grace or the exhibit explaining about them?" Just over half of the respondents (54%) indicated they did talk to someone after leaving the exhibit. About 40% did not, and 1% did not recall whether they had spoken to someone about the exhibit.

Those who indicated that they had talked to someone about the exhibit were asked the open-ended question: "Who did you talk to about Ada and Grace after your visit?" The majority talked with a family member or a person who accompanied them to the museum (Table 18). Those participants under the age of 16 were more likely to **talk to a family member regarding the exhibit** than adult participants (χ^2= 4.513, N=31, *p*=0.034). Others spoke with friends, co-workers, or someone else. Participants were encouraged to identify anyone that they spoke to, thus some mentioned more than one person.

Table 18: Responses to "*Who did you talk to about Ada and Grace after your visit?*" (n=31)

Responses	n	Percent*
A family member	23	74%
A person who went to the museum with me that day	16	52%
A friend or co-worker	14	45%
Someone else	1	3%

* Multiple responses allowed. Percentages total more than 100%.

Respondents who affirmatively responded that they talked with someone after the visit were asked the open-ended question, "What about Ada and Grace did you and the person talk about?" Over half (52%) said they talked about how the twins worked, their design, and their technology. Many (39%) had a positive comment in general to report, explaining the exhibit was enjoyable. Others described how they interacted with the twins at the exhibit (26%), or discussed the museum in general (16%). Some of the comments regarding the discussion included:

How the twins talked back to me, like they were real.
How weird it was for me to talk to a computer and it talked back.
The fact that artificial intelligence is here and people at the museum are attempting to introduce it to people from around the world.

The Twins: Visitor Attitudes

Table 19 shows the Attitude impacts and indicators for the Twins and Science Behind exhibits. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 19: Twins Attitudes High-Level Results

Impact (Shaded) and its Related Indicators	**Level of Evidence Based on the Summative Evaluation**
Children (ages 7 -14) and adults will have a positive attitude about computer science and technology.	*Achieved: 2 of 2 indicators show evidence of positive attitudes.*
Visitors will indicate they **enjoy learning about technology**.	• Participants' mean rating for their interest in learning about technology was 3.3. • The mean rating for interest in computers was 3.5. • There were no significant differences in ratings between Blended and Direct participants.
Visitors will have a **positive attitude towards virtual humans in society**, as indicated by having positive perceptions of: a) interacting with a computer, b) virtual humans having a presence and persona, c) using natural language with a computer, and d) future	• 92% of those interviewed for the study held generally positive views (although some with caution) about virtual humans in society. • Participants' median rating for speaking with the Twins was 3 out of 4; Blended rated speaking with the Twins higher than did

developments in the field of artificial intelligence.	Direct participants. • Adult participants' ratings for attitudes towards computers increase significantly after having interacted with the Twins.
Visitors will **enjoy interacting with the Virtual Humans**.	*(Subsumed by the indicator above)*

Visitors will indicate they like learning about technology (Attitudes Indicator)

Respondents were asked to rate their interest in "learning about technology" and "figuring out how computers work." Overall, respondents rated their interest in both areas fairly high (See Table 20). When additional analysis was performed, independent variables such as condition (Blended/Direct), age (adult/child), and time spent at the exhibit yielded no significant differences for either measure.

Table 20: Respondents' Ratings of their Interest in Technology and Computers (*Scale: 1="boring" and 4="exciting"*; n=174)

	How would you describe your interest in:	
	Learning about Technology is…	Figuring out how computers work is…
Mean Rating	3.3	3.5
Boring	0%	1%
Just okay	17%	12%
Pretty good	40%	37%
Exciting	43%	47%

Respondents were also asked to rate statements relative to their knowledge of technology and computers. Overall when compared to their interest ratings, respondents rated themselves as more interested than knowledgeable, with mean knowledge ratings below 3.0 (Table 21). When additional analysis was performed, independent variables such as condition (Blended/Direct), age (adult/child), and time spent at the exhibit yielded no significant differences for either measure.

Table 21: Respondents' Ratings of their Knowledge of Technology and Computers (*Scale: 1="nothing" and 4="a lot";* n=174)

	How much would you say you know about:	
	Technology	Computers
Mean Rating	2.7	2.8
Nothing	1%	1%
A little	44%	39%
Quite a bit	41%	40%
A lot!	13%	21%

Visitors will have a positive attitude towards virtual humans in society, as indicated by having positive perceptions of: a) interacting with a computer, b) virtual humans having a presence and persona, c) using natural language with a computer, and d) future developments in the field of artificial intelligence (Attitudes Indicator)

A variety of measures were used to determine the attitudes of participants towards virtual humans in society. These measures included:

- Analysis of the open-ended question "What are your feelings about virtual humans, in general?"
- Analysis of the interest rating scale question "Being able to speak with the Twins was…" which was asked of both adults and children.
- Analysis of the following attitudinal rating scale questions which were asked of adults only in a retrospective-pre/post/delayed post manner:
  - "I enjoy being able to speak to a computer as a way to interact with it."
  - "Having a computer with a personality is a good thing."
  - "In the future, there will be new and exciting innovations with smarter computers."
  - "In the future, interacting with computers will be easier."
- A cross-question analysis of the above questions to create a two-point rubric representing the indicator.

Participants' Feelings towards Virtual Humans in an Open-ended Interview Question

Near the end of the interview, participants were asked "What are your feelings about virtual humans, in general?" Piloting had demonstrated that this open-ended question could successfully elicit the feelings of interviewees relative to the status of virtual humans in society. For the summative evaluation study, responses to this question were categorized using the following rubric:

- Positive perception: Used to categorize responses such as "I think it was really interesting to talk with them and communicate with them," "They're very useful objects," and "They can sometimes be helpful with information."

- Positive but cautious perception: Used to categorize responses such as "They're okay for displays but not to take away real jobs like yours," "Nice, but creepy," and "As long as we control it, it's okay."
- Neutral perception: Used to categorize responses such as "They're alright."
- Negative perception: Used to categorize responses such as "I don't usually like them because they are slow," "It would be weird waking up with twins who are electronic in the morning asking what you want for breakfast and all," and "It's strange because it's not a real person."
- "I don't know:" A category that was created to quantify the respondents who said they did not know what their feelings towards virtual humans were.

Based upon this rubric, the majority of participants (73%) voiced a positive attitude towards virtual humans in society (Table 22).

Table 22: Responses to "*What are your feelings about virtual humans, in general?"* (n=165)

Responses	n	Percent
Positive	100	61%
Positive but cautious	21	13%
Neutral	22	13%
Negative	15	9%
"I don't know"	7	4%

The following areas of significance were found when this variable was analyzed:

- Children were more likely than adults to voice **positive views** (χ^2= 10.828, N=165, p=0.001).
- Adults were more likely than children to voice views that were **positive but cautious** (χ^2= 9.107, N=165, p=0.003) or **neutral** (χ^2= 5.331, N=165, p=0.021).
- Those who had visited Science Behind were significantly less likely to voice **positive views** than those who had not seen Science Behind (χ^2=7.665, N=165, p=0.006).
- There was a positive relationship between the interest rating for "Being able to speak with the Twins" and voicing **positive views** about virtual humans in society (Mann-Whitney U=2567.5, N=165, p=.013). In a reverse of this trend, those who were less interested in speaking with the Twins were more likely to voice **negative views** on virtual humans in society Mann-Whitney U=698.0, N=165, p=.009).
- Those voicing **neutral views** on virtual humans in society tended to rate their interest in interacting with the exhibit lower than other visitors (Mann-Whitney U=1165, N=165, p=.024).

Participants' Ratings of Being able to Speak with the Twins

A quantitative rating scale question was also used to determine if participants had a positive experience at the exhibit. Participants were asked to rate their interest in "Being able to speak with the Twins" on a four point scale, where 1 was "boring" and 4 was "exciting." As shown in Table 23, the overall rating for all participants was a median of 3, or "pretty good" on the scale. There was a statistically significant

difference in the ratings of the Blended and Direct participants, with Blended participants rating their interest higher (Mann-Whitney *U*=2577, N=174, *p*=0.000). Table 24 breaks down the frequencies for each rating.

Table 23: Interest in Speak with the Twins Rating

Being able to speak with the Twins was …	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	174	3.20	3	.745	1	4
Blended	76	3.43	4	.680	1	4
Direct	98	3.02	3	.746	1	4

Table 24: Interest in Speak with the Twins Frequencies

Rating	Overall (n=174)	Conditions	
		Blended (n=76)	Direct (n=98)
Boring	2%%	1%	2%
Just Okay	14%	7%	20%
Pretty Good	46%	40%	51%
Exciting	38%	53%	27%

Other significant findings for the rating of "Being able to speak with the Twins" include:

- Children rated speaking with the Twins **higher** than did adults (Mann-Whitney *U*=2891, N=174, *p*=0.047).
- Those who visited Science Behind rated their interest in speaking with the twins lower than did participants who did not visit Science Behind (Mann-Whitney *U*=1784, N=174, *p*=0.014). It could be that those who visited Science Behind wanted the a more in-depth experience than just speaking to the twins, leading them to go to the exhibit that explained more about how the Twins work. It is important to note that those who visited Science Behind rated the overall experience at the exhibit similarly as those who did not go to Science Behind.
- There was a positive correlation between a high interest in speaking with the Twins and **spending more time** at the exhibits (Spearman's Rho=.244, N=173, *p*=.001).
- Those participants who had a **social interaction with a staff member** tended to rate their interest in speaking with the Twins higher (Mann-Whitney *U*=2858.5, N=173, *p*=0.004).
- Those participants who had a spoke with **staff about a conceptual aspect** of the exhibit (Mann-Whitney *U*=2672.5, N=173, *p*=0.042) or an **operational aspect** of the exhibit tended to rate their interest in speaking with the Twins higher (Mann-Whitney *U*=2784.5, N=173, *p*=0.002).
- Those participants who had **positive affective interaction with a staff member** also tended to rate their interest in speaking with the Twins higher (Mann-Whitney *U*=2097, N=173, *p*=0.001).

Participants were asked this same question six weeks later in the follow-up online questionnaire; researchers compared ratings from the interview and follow-up questionnaire. Ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Adults' Ratings of Their Attitudes towards Virtual Humans

A quantitative approach was used with adults to determine if interacting with the exhibit impacted self-reported agreement with 1) "I enjoy being able to speak to a computer as a way to interact with it," 2) "Having a computer with a personality is a good thing," 3) "In the future, there will be new and exciting innovations with smarter computers," and 4) "In the future, interacting with computers will be easier." When looking at adults' ratings overall, regardless of condition, adults had a significantly higher rating for all of these measures of attitudes towards computers/virtual humans directly after their interaction with the exhibit (See Table 25).

Table 25: Adult Respondents' Ratings (Retrospective-Pre/Post) of their Attitudes towards Computers/Virtual Humans *(Scale: 1="strongly disagree," 7="strongly agree"*; n=56)

Statement	Retrospective-Pre Rating (Mean)	Post Rating (Mean)	*Z value*	*p* value
I enjoy being able to speak to a computer as a way to interact with it	4.6	5.6	-4.476	.000**
Having a computer with a personality is a good thing	4.7	5.3	-3.671	.000**
In the future, there will be new and exciting innovations with smarter computers	6.2	6.6	-2.488	.013*
In the future, interacting with computers will be easier	6.1	6.7	-3.557	.000**

Note. The Wilcoxon Signed Ranks test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

When the results were analyzed based on condition, both Blended and Direct participants showed statistically significant changes in their ratings retrospective-pre to post for the majority of the statements. This indicates that for these measures, the exhibit experience as a whole, and not the condition type, supported the changes in attitudes. The significance levels for each statement by condition are reported below:

- *Blended: I enjoy being able to speak to a computer as a way to interact with it.* Significant change in mean rating retrospective-pre to post (Z=-2.911, N=21, p=.004).
- *Direct: I enjoy being able to speak to a computer as a way to interact with it.* Significant change in mean rating retrospective-pre to post (Z=-3.561, N=35, p=.000).
- *Blended: Having a computer with a personality is a good thing.* Significant change in mean rating retrospective-pre to post (Z=-2.041, N=21, p=.041).
- *Direct: Having a computer with a personality is a good thing*. Significant change in mean rating retrospective-pre to post (Z=-3.097, N=35, p=.002).
- *Blended: In the future, there will be new and exciting innovations with smarter computers*. **No significant change** in mean rating retrospective-pre to post (Z=-.577, N=21, p=.564).

- *Direct: In the future, there will be new and exciting innovations with smarter computers.* Significant change in mean rating retrospective-pre to post (*Z*=-2.410, N=35, *p*=.016).
- *Blended: In the future, interacting with computers will be easier.* Significant change in mean rating retrospective-pre to post (*Z*=-2.070, N=21, *p*=.038).
- *Direct: In the future, interacting with computers will be easier.* Significant change in mean rating retrospective-pre to post (*Z*=-2.966, N=35, *p*=.003).

Adult participants in the follow-up questionnaire were asked to rate the statements again using the same scale. When comparing the onsite ratings with the online questionnaire ratings (i.e. post to delayed post), there were no significant differences in how three of the four statements were rated ("I enjoy being able to speak to a computer as a way to interact with it;" "Having a computer with a personality is a good thing;" and "In the future, there will be new and exciting innovations with smarter computers.") In other words, respondents' attitudes held constant in the six weeks after the museum visit for these statements. This could be interpreted as the gains that were made directly after interacting with the exhibit were maintained overtime. Another possibility is that these results are a "practice effect" from answering the same question multiple times. Also possible is that not enough time had passed for a decay of the effects to have occurred. A questionnaire conducted six months after having used the exhibit may yield different results. For the last statement, "In the future interacting with computers will be easier," there was a significant difference post to delayed post the ratings, with lower delayed post ratings (*Z*=-1.983, N=14, *p*=.047). This indicates a "decay" in the attitudes towards the ease of interacting with computers in the future, with attitudes returning to the retrospective-pre levels.

Overarching Rubric for Participants' Attitudes towards Virtual Humans in Society

An overarching variable was created as a gestalt measure of participants' attitudes towards virtual humans in society. The following rubric was created to address the indicator as a whole:

- Participant DOES display a positive attitude towards virtual humans in society: Included participants who 1) voiced either a "positive" or "positive with caution" to the open-ended question, 2) rated "Being able to speak with the Twins" as a 3 or 4, OR 3) had a rating of 5 or higher to any of the post rating questions (adults only). A positive attitude for any of these questions included the participant in this category.
- Participant does NOT display a positive attitude towards virtual humans in society: Included participants who 1) did not voice either a "positive" or "positive with caution" to the open-ended question, 2) rated "Being able to speak with the Twins" as a 1 or 2, and 3) had a rating of 4 or lower to all of the post rating questions (adults only). Participants had to meet all of these criteria to be included in this category.

Based upon this rubric, the vast majority (92%) of those interviewed for the study held generally positive views (although some with caution) about virtual humans in society. There were no statistically significant findings when this variable was analyzed by condition (Blended/Direct), age (adult/children), time spent at the exhibit or visitation to Science Behind.

The Twins: Visitor Awareness

Table 26 shows the Awareness impacts and indicators for the Twins and Science Behind exhibits. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 26: Twins Awareness High-Level Results

Impact (Shaded) and its Related Indicators	Level of Evidence Based on the Summative Evaluation
Children (ages 7 -14) and adults will increase their awareness about computer science and technology.	*Achieved: 4 of 5 indicators show evidence of an increase in awareness.*
Visitors will increase their **awareness of what a virtual human is**.	• Adult participants' awareness rating for understanding what a virtual human was significantly higher after having interacted with the exhibit.
Visitors will increase their **awareness of the role of women as role models in computer science**.	• Adult participants' awareness rating for the role of women in computer science was significantly higher after having interacted with the exhibit.
Visitors will be able to **describe Ada & Grace as a computer that acts like a human**.	• 94% of participants described the Twins in both human and computer-like terms. • Participants were able to name attributes of the Twins that were human-like and computer-like.
Visitors will recognize at least one of the following as **characteristics of the Twins**: a) they interact through speech, b) they are able to respond, c) their responses are dependent on what is asked of them, d) they have non-verbal behaviors.	• 93% of participants used at least one of the categories listed in the indicator as they talked about the Twins and their experience at the exhibit.
Visitors will recognize Ada and Grace as **relating to the objectives of Cahners Computer Place**, with the following indicators of awareness: 1) they highlight the same subjects as Cahners (computers, communications, robots), 2) they are "guides" to the space, directing visitors to other exhibits in Cahners.	• 39% of participants were aware of at least one of these aspects of the connection between the Twins and the objectives of Cahners Computer Place.

Visitors will increase their awareness of what a virtual human is (Awareness Indicator)

A quantitative approach was used with adults to determine if interacting with the exhibit impacted self-reported agreement with the statement "I understand what a virtual human is." When looking at adults' ratings overall, regardless of condition, adults had a significantly higher rating of for this measure of awareness of a virtual human directly after their interaction with the exhibit:

- The mean retrospective-pre rating of the statement was 4.6 compared to 5.9 as the post rating, with 7 as "strongly agree" (Z=-4.704, N=56, p=.000).

- Adults in both the Direct and Blended conditions had significantly higher ratings after their interactions with the exhibit (Direct: *Z*=-4.092, N=35, *p*=.000; Blended: *Z*=-2.504, N=21, *p*=.012) indicating that the exhibit as a whole regardless of condition improved visitors awareness of what a virtual human is.

Adult participants in the follow-up questionnaire that occurred six weeks later were asked to rate the same statement using the same scale. When comparing the onsite ratings with the online questionnaire ratings (i.e. post to delayed post), there were no significant differences in the ratings for the statement (*Z*=-.791, N=14, *p*=.429). In other words, respondents' awareness of what a virtual human is held constant in the six weeks after the museum visit. This could be interpreted as the gains that were made directly after interacting with the exhibit were maintained overtime. Another possibility is that these results are a "practice effect" from answering the same question multiple times. Also possible is that not enough time had passed for a decay of the effects to have occurred. A questionnaire conducted six months after having used the exhibit may yield different results. There were no differences between the ratings of the treatment and control groups in how this statement was rated in the delayed post.

Visitors will increase their awareness of the role of women as role models in computer science (Awareness Indicator)

A quantitative approach was used with adults to determine if interacting with the exhibit impacted self-reported agreement with the statement "Women have made important contributions in the field of computer science." When looking at adults' ratings overall, regardless of condition, adults had a significantly higher rating of for this measure of awareness of women's roles directly after their interaction with the exhibit:

- The mean retrospective-pre rating of the statement was 5.8 compared to 6.3 as the post rating (*Z*=-3.335, N=56, *p*=.001).
- Adults in the Direct condition only had significantly higher ratings after their interactions with the exhibit (*Z*=-3.090, N=35, *p*=.002). Participants interacting with the Blended condition did not show a significant change in rating retrospective-pre to post (*Z*=-1.342, N=21, *p*=.180). This indicates that visitors in the Blended condition did not have an increase in awareness of the roles of women in computer science as a result of interacting with the exhibit.

Adult participants in the follow-up questionnaire that occurred six weeks later were asked to rate the same statement using the same scale. When comparing the onsite ratings with the online questionnaire ratings (i.e. post to delayed post), the delayed ratings were significantly lower than the post ratings (*Z*=-2.308, N=14, *p*=.021). This indicates a "decay" in the level of awareness over the six weeks with awareness levels returning to retrospective-pre levels.

Visitors will be able to describe Ada & Grace as a computer that acts like a human (Awareness Indicator)

In order to determine whether visitors were able to describe the Twins, Ada & Grace, as a computer that acts like a human, visitors were asked the following, open-ended questions: 1) "How would you describe the Twins, Ada and Grace, to someone else?"; 2) "In what ways do you think Ada and Grace are the same as a real person?"; and 3) "In what ways are Ada and Grace different from a real person?" In addition to the analysis of these open-ended questions, a quantitative variable was created as an

indicator of this indicator. The results of the open-ended questions and the indicator variable are reported below.

Participants' Description of the Twins

In order see whether, unprompted, visitors would describe the Twins as a computer that acts like a human, they were first asked how they would describe the Twins, Ada and Grace to someone else. Responses to this open-ended question were coded and quantified by ILI researchers into three broad categories: 1) general reactions and attitudes towards the twins (including positive and/or negative reactions); 2) perceptions of the technological advancement of the twins (including perceptions of limited and/or advanced technologies); and 3) awareness of the twins as being human-like, non-human-like, and/or both. A forth category was created to include responses where visitors said they did not know how to describe the Twins.

When asked to describe the Twins, just over two-thirds (76%) of respondents' top-of-mind responses centered on the Twins human and/or non-human-like attributes, nearly one-fifth of visitors (19%) described the Twins more broadly in either a positive or negative light, 12% described the twins as being either technologically advanced or limited, and 7% said they did not know how to describe the Twins (See Table 27).

Table 27: Responses to "*How would you describe the Twins, Ada & Grace, to someone else*" (n=180)

Responses	n	Percent*
Human / Non-Human-like Attributes	136	76%
Human-like	62	34%
Non-human-like	29	16%
Both human and non-human-like	45	25%
General Reactions	35	19%
Positive	31	17%
Negative	5	3%
Technological Advancement	22	12%
Advanced technology	11	6%
Limited abilities	11	6%
Don't know/unsure	13	7%

* Multiple responses allowed. Percentages total more than 100%.

Human / Non-Human-like Attributes: Within this category, most respondents described the Twins as having human-like attributes (34%), followed by non-human-like characteristics (16%), or both types of attributes (25%). Visitors who described the Twins as having human-like characteristics referred to the Twin's as being similar to humans *cognitively* (in that they have ability to communicate and "think" about responses), *physically* (in that they look like and move like humans), *socially* (in that they have personalities and display social niceties), *emotionally* (in that they display emotions and humor), and

circumstantially (in that they are Twins, scientists, and so on). Some examples of these responses include:

They both have facial expressions, can smile, and be mad.
They look alike and they move their arms.
They act like sisters – they bicker a lot.
You can have a real conversation with them.
They can understand what you are saying.

Visitors who described the twins as not being similar to humans referred to the computer-based aspects of the Twins such as being projected on a screen, not sounding like a human (both in voice and content) or being run on a computer. Visitors in this category also called the Twins "robots" or "avatars."

Visitors who cited both human and non-human traits as they responded to the question combined both types of qualities as they explained the Twins. Typical responses include:

They can talk, but they aren't real.
They are virtual humans and you can talk to them.
They run on a computer and act like humans. You say things and they respond back but they are not humans, they are computer programs that act like humans.
Robotic people.

General Reactions: Around a fifth of respondents described the Twins broadly, either in a positive (17%) or negative (3%) light. Visitors who reacted positively to the Twins described them as "awesome," "cool," and "interesting." Visitors who had a less positive experience and/or perception of the Twins described them as "uncanny," and "weird."

Technological Advancement: A few visitors described the Twins as being either technologically advanced (6%), or limited in their current state (6%). Visitors who perceived the Twins as advanced referred to them as being "knowledgeable," "smarter than normal humans," and "intelligent." Visitors who perceived the Twins as being somewhat limited, mostly referred to the Twin's inability to answer more than a set number of questions and/or understand certain voices and accents.

When each category of response was compared by type of interaction (Blended or Direct) and age group (adults and children), no significant differences were discovered (Chi Square test). When each category of response was compared by time and knowledge and interest of computers and technology, one significant difference was discovered. Visitors who reported having greater knowledge of technology were more likely to describe the Twins as having only limited technological abilities (Mann-Whitney U=574.5, N=177, p=0.26).

Participants' Views of How the Twins are Similar to Humans

In a more direct approach, visitors were asked to discuss the ways in which Ada and Grace are the same as a real person. Responses to this question were coded and quantified by ILI researchers into five categories: 1) Talk like humans, or are socially similar; 2) Look like humans, or are physically similar; 3) Think like humans, or are cognitively similar; 4) Act like humans, or are emotionally similar; and 5) the Twins are not at all the same as a real person. Two additional categories were created to capture

responses where visitors said they were either unsure how the Twins are the same, or they provided a response that was un-codeable or was un-related to the question.

Over half of visitors (59%) said the Twins are socially similar to real humans in that they can talk and have conversations; almost half (49%) said they are physically similar in that they look like humans; almost a quarter (24%) said they are cognitively similar in that they think like humans; and 13% said they are emotionally similar in that they act or react like real humans. Only a small percentage of visitors (7%) said that the Twins are not at all like a real person. (See Table 28)

Table 28: Responses to "*In what ways do you think Ada and Grace are the same as a real person"* (n=179)

Responses	n	Percent*
Talk like humans (Socially similar)	105	59%
Look like humans (Physically similar)	87	49%
Think like humans (Cognitively similar)	43	24%
Act/React like humans (Emotionally similar)	23	13%
Not the same as a real person	12	7%
Don't know/unsure	2	1%
Other / un-codeable response	7	4%

* Multiple responses allowed. Percentages total more than 100%.

Over half of visitors discussed the Twins' ability to talk like humans. These visitors mostly described how the Twins "had voices" and were able to "speak" and "answer questions." An almost equal percentage of visitors described how the Twins looked similar to real humans, pointing out aspects such as their "body shapes," "facial expressions," "body movements," and "outfits." Some visitors described the cognitive abilities of the Twins, noting how they were not only able to speak, but were able to listen to and understand visitors' questions. These visitors said things like:

They listen and then are able to answer your questions.
They think about what they can answer.
They can understand you.

Fewer visitors described the Twins as being emotionally similar to real humans. Those who did perceive emotionally similar characteristics described the Twins as having "personalities" and "feelings." Visitors who did not find any similarities between the Twins and real humans, tended to point out differences, saying things like: "They don't understand what you say," and "A real person can answer a lot of questions to talk."

When each category of response was compared by type of interaction (Blended or Direct), no significant differences were found (Chi Square test). When comparing between age groups (adults and children), significant differences were discovered.

- Children were significantly more likely to say that the Twins **talk like real humans** (socially similar), than were adults (χ^2= 4.524, N=175, p=0.033).

- Adults, on the other hand, were significantly more likely to note **emotional similarities** between the Twins and real humans (χ^2= 4.544, N=175, *p*=0.033).

Significant differences were also discovered when each category of response was compared by time and knowledge and interest in computers and technology.

- Visitors who spent longer amounts of time at the exhibit, were significantly more likely to describe the Twins as being able to **talk like real humans** (Mann-Whitney *U*=3186.5, N=179, *p*=0.041).
- Looking at knowledge and interest, visitors who reported having lower interest in learning about technology were significantly more likely to say that the Twins **are not the same as real humans** (Mann-Whitney *U*=616.0, N=176, *p*=0.20).

Participants' Views of How the Twins are Different from Humans

As a reverse to the above question, visitors were asked to describe the ways in which Ada and Grace are different from a real person. Responses to this question were coded and quantified by ILI researchers into four overarching categories: 1) They are computers (no further elaboration); 2) They lack feelings and emotions; 3) They lack mobility and other physical characteristics; and 4) They lack full communication skills and the ability to learn, think, and understand. Two additional categories were created to capture responses where visitors said they were either unsure how the Twins are different from a real person, or they provided a response that was un-codeable or was un-related to the question.

Close to half of visitors said that Ada and Grace were different from real humans in that they lacked certain physical characteristics and mobility (46%) or that they lacked full communication skills and/or the ability to think, learn, and understand (46%). One fifth of visitors (20%) did not elaborate in great detail about the differences, but did indicate an awareness of Ada and Grace being "computers." Only a small percentage of visitors (6%) described the Twins as being different from humans emotionally, in that they lacked the ability to have and express feelings. A few people said they did not know how the Twins were different (3%) and others provided responses that were either un-codeable or fell into some other category (5%). (See Table 29)

Table 29: Responses to "*In what ways are Ada and Grace different from a real person"* (n=179)

Responses	n	Percent*
Lack mobility and other physical characteristics	82	46%
Lack full communication skills and the ability to learn, think, understand	82	46%
It's a computer (general)	36	20%
Lack feelings and emotions	11	6%
Don't know/unsure	6	3%
Other / un-codeable response	9	5%

* Multiple responses allowed. Percentages total more than 100.

Visitors who mentioned physical limitations gave responses such as:

They are just animated. They can't really move.
They can't come out and shake hands with you. They can't do things with you.
They can't climb mountains.
They can't change their clothes.

Visitors commonly spoke of the ways the Twins are limited in how they communicate. For many of these visitors, limitations in communication abilities were related underlying cognitive limitations; therefore, these concepts were grouped together. Typical responses in this category included:

They have the same speech patterns for every sentence.
They will never duplicate neural processing. They just answer to canned responses -- if you pose new questions, they can't answer, [whereas] humans could. Even with different syntax and different accents, [humans] would understand.
A normal person would hear you clearer. A normal person would know what you're talking about. They're virtual computers.
They don't think like humans – they're programmed to do something.

Some visitors explained that the biggest difference between Ada and Grace and real humans is simply that the Twins are computers. These visitors spoke very broadly, saying things like "they are computers," or "they are virtual humans and we are real humans."

Finally, a small percentage of visitors indicated that a difference between the Twins and humans, is that Ada and Grace appeared to lack feelings and emotions[3]. These visitors said things like:

They couldn't take my joke.
I suppose if you believe in soul and so on they don't have any soul.
They don't have feelings, they're just robots.

When each category of response was compared by type of interaction (Blended or Direct), no significant differences were found (Chi Square test). When comparing between response categories and age group (adults and children), significant differences were discovered.

- Children were significantly more likely to provide a general response, saying that the Twins are different because **they are computers** (χ^2= 12.472, N=176, p=0.000) and/or were different physically (χ^2= 4.979, N=176, p=0.026) than were adults.
- Adults, on the other hand, were significantly more likely to note **cognitive** (χ^2= 16.503. N=176, p=0.000) and/or **emotional differences** (χ^2= 5.184, N=176, p=0.023) between the Twins and humans than were children.

Significant differences were also discovered when each category of response was compared by time spent and knowledge and interest in computers and technology.

[3] The team at ICT asked researchers at ILI to investigate the level of overlap between the categories of a) "lack feelings and emotions" and b) "lack communication skills/ability to think, learn, and understand." Of the 11 individuals who mentioned emotional limitations of the Twins, 5 also had a response coded as lacking communication/cognitive abilities.

- Visitors who spent longer amounts of time at the exhibit, were significantly more likely to describe the Twins as **lacking emotional capacity** (Mann-Whitney *U*=580.000, N=180, *p*=0.037) and/or as being computers (Mann-Whitney *U*=1898.500, N=180, *p*=0.013).
- Looking at knowledge and interest, visitors who reported having higher interest in figuring out how computers work were significantly more likely to say that the Twins **lacked physical characteristics and mobility** (Mann-Whitney *U*=3188.500, N=177, *p*=0.026).

Overarching Rubric for Participants' Awareness of the Twins as a Computer that Acts like a Human

To create a variable for the analysis of the indicator, all the responses given by a participant throughout their entire interview were reviewed. A rubric was used to categorize participants' awareness of the Twins as computers that act like humans. The rubric included two categories:

- Participant DOES describe the Twins as a computer that acts like a human: Participants in this category made reference to human-related aspects of the Twins AND to non-human/computer-related aspects of the Twins. Human-related aspects included that the Twins interact through speech, are able to respond to questions asked of them, have non-verbal behaviors, look like humans, think like humans, or react like humans. Non-human/computer-related aspects include that the Twins are computers, programmed, AI, "not real," or need the support of a screen, microphone, and speakers to interact with visitors.
- Participant does NOT describe the Twins as a computer that acts like a human: Participants in this category did not make reference to human-related AND non-human/computer-related aspects. This category included visitors who referenced only one of these and not the other.

Based on this rubric, the vast majority (94%) of those interviewed described the Twins in both human and computer-like terms. Analysis of the independent variables (condition, age, time spent at the exhibit, and visitation to Science Behind) demonstrated no significant differences.

Visitors will recognize at least one of the following as characteristics of the Twins: a) they interact through speech, b) they are able to respond, c) their responses are dependent on what is asked of them, d) they have non-verbal behaviors (Awareness Indicator)

To create a variable for the analysis of this indicator, all the responses given by a participant throughout their entire interview were reviewed. A rubric was used to categorize participants' awareness of these characteristics of the Twins. The rubric included two categories:

- Participant DOES recognize at least one of the categories in the indicator: Participants in this category made reference to at least one of the categories listed in the indicator: a) the Twins interact through speech, b) they are able to respond, c) their responses are dependent on what is asked of them, AND/OR d) they have non-verbal behaviors.
- Participant does NOT recognize any of the categories in the indicator: Participants in this category did not make reference any of the above categories.

Based on this rubric, the vast majority (93%) of those interviewed used at least one of the categories as they talked about the Twins and their experience at the exhibit. Analysis of the independent variables

(condition, age, time spent at the exhibit, and visitation to Science Behind) demonstrated no significant differences.

Visitors will recognize Ada and Grace as relating to the objectives of Cahners Computer Place, with the following indicators of awareness: 1) they highlight the same subjects as Cahners (computers, communications, robots), 2) they are "guides" to the space, directing visitors to other exhibits in Cahners (Awareness Indicator)

To determine if visitors understood the relationship between the Twins and gallery in which they were found (Cahners Computer Place), participants in the interviews were asked one yes/no question and one open-ended question. These were analyzed and the open-ended question was used to create a quantitative awareness scale to represent the indicator.

When asked "Did interacting with Ada and Grace help you to better understand what activities there are in this gallery?", participants were divided in their responses; 47% replied "yes," the Twins did help, 43% said "no," and another 10% said they were not sure if the twins were helpful in understanding the gallery. The following areas of statistical significance were found for this question:

- Adults were less likely than children to say that interacting with the Twins helped them to **better understand** what was in the gallery (χ^2= 8.835, N=172, p=0.012). This may be related to adults' greater awareness of and familiarity with the museum or gallery in general.
- Those participants who indicated that the Twins did help them to **better understand** what was in the gallery had spent significantly longer at the exhibit (Mann-Whitney U=2200.5, N=156, p=0.003).
- Note that condition (Blended/Direct) was not statistically significant; those under the Blended condition were not more likely to say "yes" than those under the Direct interaction condition.

Those who indicated that the Twins were helpful were then asked "How did they [the Twins] help?" The open-ended responses to this question were categorized into a three-point rubric, described below:

- No recognition of the connections (0 points): Participants who 1) replied that the Twins did not help them to understand the gallery, 2) were unsure how the Twins might have helped, OR 3) in the open-ended question did not mention either of the topics in the indicators were considered to have made no connections between the goal of the Twins and the rest of Cahners.
- Recognizes one connection in the indicator (1 point): Participants who mentioned one of the two topical connections between the Twins and Cahners in their open-ended response were coded into this category. The two topics were: 1) The Twins highlight the same subjects as Cahners (computers, communications, robots) and 2) they are "guides" to the space, directing visitors to other exhibits in Cahners[4].

[4] A few participants in the interview indicated elsewhere in the interview (typically in the answer to "What would you say was the main idea of the exhibit?" or "How would you describe the Twins, Ada and Grace, to someone else?") that the Twins were guides. For these participants, this connection was counted even though it had not occurred in answer to the questions described above.

- Recognizes both connections in the indicator (2 points): Participants who mentioned both of the two topical connections (as described above) between the Twins and Cahners were coded into this category.

Based upon this scales the average number of connections made by participants was 0.4 connections per participant. Looking at the percent of responses to category of the rubric 62% made no connections, 36% made one connection, and 3% made both connections relative to the Twins and Cahners. For the purposes of the indicator, 39% of those interviewed were aware of the connection between the Twins and the objectives of Cahners Computer Place. The following areas of significance were found when this variable was analyzed:

- Time spent the exhibit was positively correlated with a higher mean **number of connections** (Spearman's Ro=0.161, N=178, *p*=0.031).
- Note that condition (Blended/Direct) was not statistically significant; those under the Blended condition were not more likely to make connections between the Twins and the goals of the gallery than those under the Direct interaction condition.

The Twins: Visitor Knowledge

Table 30 shows the Knowledge impacts and indicators for the Twins and Science Behind exhibits. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 30: Twins Knowledge High-Level Results

Impact (Shaded) and its Related Indicators	**Level of Evidence Based on the Summative Evaluation**
Children (ages 7 -14) and adults will increase their knowledge about computer science and technology.	*Achieved: 3 of 3 indicators show evidence of increased knowledge.*
Visitors will be able to **name at least one aspect of what makes up virtual humans**, such as: a) communications technology, b) artificial intelligence, c) natural language, d) animation/graphics, or e) nonverbal behavior.	• 73% of all participants mentioned two or more aspects listed in the indicator. • The most commonly mentioned aspect was natural language, mentioned by 86% of participants.
Visitors will be able to **discuss at least one technology that is needed to build a virtual human**.	• 64% of onsite participants named at least one technology needed to build a virtual human; this rose to 90% in the follow-up six weeks later. • The most commonly mentioned technology was computers (43%) followed by voice or speech recognition-related technologies (26%).
Visitors will **learn at least one new idea related to a STEM domain** supported by the Twins.	• 84% of participants gained at least one additional understanding about STEM domains related to the Twins. • 59% of participants indicated they learned something new about computers or

	technology from interacting with the exhibit. • The most common new learning was related to the technological capabilities and advancements evidenced by the Twins.

Visitors will be able to name at least one aspect of what makes up virtual humans, such as: a) communications technology, b) artificial intelligence, c) natural language, d) animation/graphic, or e) nonverbal behavior (Knowledge Indicator)

To determine whether participants who interacted with the exhibit recognized the various aspects of computer science needed to create a virtual human, a series of variables were created. First an overview of all the responses in a participant's interview was used to note the presence or absence of each aspect in the indicator. Then an overarching variable was created for analysis in response to the indicator itself.

Participants' interviews were coded for the presence or absence of each aspect in the indicator (Table 31); this analysis revealed that 86% of all participants gave responses related to the Twins' use of natural language. It is important to note that most visitors did not use the phrase "natural language," instead they described aspects that could be coded as the use of natural language, such as speech as an input and output for the Twins, and the idea that the Twins interact through and are able to respond through speech. Participants who mentioned voice modeling or speech recognition were also captured in this category.

The second most common aspect mentioned was the usage of animation or graphics, with 62% of participants giving a response in this category during the interview. This included noting the technology that was needed to display the twins (a screen), the techniques used to create the animation (face modeling), and how the animation of the Twins appears ("life-like").

Table 31: Participants' mentions of five aspects of computer science that make up a virtual human (n=180)

Responses	n	Percent*
Natural language	154	86%
Animation/Graphics	112	62%
Intelligent reasoning/Artificial intelligence	52	28%
Non-verbal behaviors	38	21%
Communications technology	23	13%

* Multiple responses allowed. Percentages total more than 100.

An analysis of these categories by the independent variables revealed the following areas of statistical significance:

- Adults were more likely than children to mention both **artificial intelligence** (χ^2= 19.836, N=176, p=0.000) and **non-verbal behaviors** (χ^2= 6.336, N=176, p=0.012).
- It is important to note that there were no differences based on condition (Blended/Direct) in which aspects of virtual humans participants mentioned during their interviews.

An overarching variable was created using the following three-point rubric:

- Participant does NOT name any of the aspects (0 points): None of the five aspects in the indicator were mentioned by the participant in the course of their interview.
- Participant names at least ONE of the aspects (1 point): One of the five aspects in the indicator were mentioned by the participant in the course of their interview.
- Participant names TWO or more of the aspects (2 points): Two or more of the five aspects in the indicator were mentioned by the participant in the course of their interview.

Using this variable, participants' the mean knowledge level of aspects needed to create a virtual human was 1.7 on the scale. Nearly three-fourths (73%) of all participants mentioned two or more aspects, 24% mentioned one, and just 3% did not mention any of the aspects during their interview. Analysis by the independent variables (condition, age, time spent at the exhibit, and visitation to Science Behind) revealed no significant differences. This indicates that the exhibit was equally successful for visitors across a variety of factors.

Visitors will be able to discuss at least one technology that is needed to build a virtual human (Knowledge Indicator)

A series of two open-end questions and a yes/no question were used to better understand how participants in the interview perceived the technology needed to make the Twins. The answers to these questions were used to create a quantitative variable for further analysis of the indicator.

Open-ended Responses for Technologies Needed to Make the Twins Work

Participants in the interviews were asked what technologies they thought were needed to make Ada and Grace work. As seen in Table 32, the most commonly mentioned response (43%) was the idea that the Twins need a computer to work: "Most definitely a computer" responded one participant with a typical response in this category. Other participants were more general still, mentioning only that the Twins used "something high tech" or lots of electricity, machines and wires (12%).

Participants who mentioned specific and more sophisticated technology used in the Twins were most likely to mention aspects of voice or speech recognition (26%); this included mentions of fairly obvious devices like a microphone to speak into as well as more sophisticated responses such as "voice recognition software". Another common response was programming or a "database of knowledge" from which to draw on (21%). Another technology that was mentioned less often was the graphics or the display needed to visualize the Twins; this was mentioned by only 11% of those interviewed. It should be noted that responses that mentioned both "a computer" as well as a more detailed response such as "voice recognition software" were coded twice, once in each category.

Table 32: Responses to "*What technologies do you think are needed to make Ada and Grace work?"* (n=174)

Responses	n	Percent*
Computers & Processing (non-specific)	75	43%
Voice / Speech Recognition	45	26%
Programming or knowledge-base	36	21%
General technology (machines, electricity)	21	12%
Graphics / Display	20	11%
Computer Scientists	7	4%
"I don't know"	39	22%
Other/un-codeable	14	8%

*Multiple responses allowed. Percentages total more than 100.

Additional analysis using the independent variables yielded the following areas of statistical significance:

- Participants in the Blended condition were significantly more likely to respond that the Twins needed **computers and processing** (χ^2= 6.653, N=174, p=0.010).
- Adults were significantly more likely to mention **voice/speech recognition** (χ^2= 8.502, N=172, p=0.004) and **programming/knowledge-base** (χ^2= 14.405, N=172, p=0.000) than were children. Interestingly, adults were just as likely as children to indicated that they "didn't know" what technology was needed (χ^2= 3.382, N=172, p=0.066).
- Significantly more time was spent by those who indicated **computers and processing** (Mann-Whitney U=2968.5, N=173, p=0.030) was used by the Twins than those who did not give this response.
- Those who answered that the Twins needed **computers and processing** rated their knowledge of technology significantly higher (Mann-Whitney U=2687.5, N=172, p=0.002), as did those who responded that the Twins need **programming/knowledge-base** (Mann-Whitney U=1711.5, N=172, p=0.005). Those responding that they **"didn't know"** what technologies were needed had statistically significantly lower ratings for their knowledge of technology (Mann-Whitney U=1844, N=172, p=0.003).

Understanding Scale for Technologies Needed to Make the Twins Work

The majority of those interviewed (66%) were able to name a technology (i.e. gave an answer other than "I don't know", "other" or "computer scientists"). However, the responses showed varying levels of understanding of the technologies needed to make the Twins work. Therefore, a rubric was created to quantify the understanding of the participants. This three-point rubric is described below:

- No understanding (0 points): Participants who replied "I don't know" or mentioned a general technology (like electricity) were considered to have demonstrated no understanding of how the Twins work.
- Basic understanding (1 point): Participants who replied by saying the Twins needed a computer or something with processing power were considered to have demonstrated basic understating of the technology needed to make the Twins work.

- Advanced understanding (2 points): Participants mentioning voice and speech recognition, programming, or graphics and display were considered to have demonstrated an advanced understanding of the technologies need.

Based upon this scale, the average level of understanding was a 1.1, or a "basic understanding". However, when looking at the percents of these categories, a different picture emerges: 42% of those interviewed were categorized as having "advanced understanding," 22% as having "basic understanding", and 36% as "No understanding" of the technology needed.

- Analysis of this variable by age reveals that **adults and children** had significantly different level of understanding: the mean for adults was 1.3 as compared to 0.9 for children (Mann-Whitney U=2523, N=172, p=0.003).
- There were no significant differences based on condition (Direct/Blended), having visited Science Behind, or time spent at the exhibit.

In an attempt to learn more about the level of technological sophistication of visitors, the interview required the data collector to ask an additional question of those who had given vague answers (such as "I don't know" or "a computer") to the open-ended technology question. These visitors were asked if they thought the Twins needed a computer to work (a yes/no question), and if yes, "how do you know if needs a computer." The answers to these two questions were used to perform further analysis on the sub-set of visitors who had been categorized as having "no understanding."

A total of 55 respondents who were categorized as having "no understanding" were asked the above questions. Of these respondents, 48 (87%) said "yes," the Twins did need a computer to work; 7 respondents (13%) said "no". Although the sample size was not large enough to merit further statistical analysis, all 7 individuals who said "no" were children. Those who said "yes" the Twins needed a computer to work primarily drew on their prior or incoming knowledge to come to this determination (78%). Very few (12%) indicated that they had learned the Twins needed a computer to work by interacting with the exhibit or the staff.

Follow-up Online Questionnaire Results for Technologies Need to Build a Virtual Human

Participants in the follow-up study were asked the same open-ended question about technologies needed to make the Twins work. As seen in Table 33 below, the most commonly mentioned technologies in the onsite questionnaire were also the most commonly mentioned in the follow-up. However, voice or speech recognition was mentioned by 50% of follow-up respondents compared to only 26% of onsite respondents. In a departure from the on-site responses to this question, follow-up participants also tended to talk about the limitations or areas of possible improvements to the Twins when answering this question; 19% of all follow-up respondents had responses of this type.

Table 33: Responses to the follow-up question "*What technologies do you think are needed to make Ada and Grace work*?*"* (n=54)

Responses	n	Percent*
Computers & Processing (non-specific)	27	50%
Voice / Speech Recognition	27	50%
Programming or knowledge-base	14	26%
Graphics / Display	14	26%
General technology (machines, electricity)	7	13%
Computer Scientists	1	2%
"I don't know"	3	6%
Other/un-codeable	4	7%

* Multiple responses allowed. Percentages total more than 100.

Using the same three-point rubric described above, follow-up participants were categorized as having a) no understanding, b) basic understanding, or c) advanced understanding of the technologies needed to create virtual humans. The mean level of understanding for follow-up participants was 1.7; this was a statistically significant change from the onsite level of understanding (Wilcoxon Signed Ranks Test: Z= -3.323, N=51, *p*=0.001). When looking at the percents of these categories, 75% of those in the follow-up sample were categorized as having "advanced understanding," 15% as having "basic understanding", and 10% as "No understanding" of the technology needed.

Visitors will learn at least one new idea related to a STEM domain supported by the Twins (Knowledge Indicator)[5]

Two open-end questions were asked to gain an understanding of how the exhibit supported STEM learning for visitors. The responses to these questions were then quantified and a new variable was created to address the indicator.

Main Idea of the Exhibit

Participants in the interviews were asked what they felt the main idea of the exhibit was. The majority felt the exhibit was in place to support "learning" in a broad sense, including the promotion of new knowledge, awareness and attitudes (Table 34). Nearly half (45%) felt that the exhibit was created to support an increase in knowledge, awareness, or positive attitudes about computers and/or technology. Responses of this type included "To get you to know more about computers and how they work" and

[5] This indicator was originally "Visitors will learn at least one new idea related to the STEM domain they talked to the Twins about." It was modified due to the complexities of analysis using the actual utterances of participants. Also the updated indicator took into account that new learning could be about the Twins themselves and their capabilities and not limited to topics that a visitor talked about with the Twins.

"To show people how computers have advanced from a bunch of circuit boards that took up a whole warehouse." Adults were more likely than children to give this response (χ^2= 4.179, N=172, p=0.041); there were no differences based on condition (Blended/Direct), time spent, or visitation to Science Behind.

A third (33%) of participants felt the exhibit supported an increase in "learning" about virtual humans in general or the Twins specifically; a representative example from this category is "To give people an understanding of how a VH works." Another 10% of participants felt the main purpose of the exhibit was for the Twins to act as tour guides to Cahners or the museums. Finally, some visitors (10%) felt that the exhibit supported interest and curiosity in general or learning about non-science subjects (such as public speaking) and others felt it supported "learning" about science in general (4%). For those categories with a sufficient number of responses, additional analyses were conducted; none yielded significant differences.

Table 34: Responses to "What would you say was the main idea of the exhibit?" (n=174)

Responses	n	Percent*
Increase knowledge, awareness, positive attitudes about *computers & technology*	78	45%
Increase knowledge, awareness, positive attitudes about *virtual humans or the Twins specifically*	58	33%
For the Twins to act as tour guides	17	10%
Stimulate curiosity and interest *generally* or about *non-science subjects*	17	10%
Increase knowledge, awareness, positive attitudes about *science in general*	7	4%
"I don't know"	10	6%

* Multiple responses allowed. Percentages total more than 100.

Learning about Technology and Computers at the Exhibit

Participants were asked two questions relative to what they may have learned at the exhibit: "Did you learn anything new about technology or computers from interacting with the exhibit?" and if yes, "what did you learn?" More than half of those interviewed (59%) said that yes, they did learn something new about computers or technology; 35% said they did not and 6% were unsure if they had learned anything new. Additional analysis of the data revealed the following trends:

- Those in the Blended condition were much more likely to indicated they **had learned something new** (72% compared to 49% in the Direct condition; χ^2= 10.173, N=178, p=0.006).
- Those who answered that they **had learned something new** spent significantly more time at the exhibit than those who had answered "no" (Mann-Whitney U=2203, N=167, p=0.000).
- Those who indicated they **did not learn anything new** had a significantly higher self-reported rating of knowledge of technology than those who said they had learned something new (Mann-Whitney U=2646, N=166, p=0.042).

- Interestingly, adults were just as likely as children to say they had learned something new.

Those participants who indicated they had learned something new were then asked what they had learned. The most common response, given by 63% of those who had learned something new, was related to the technological capabilities and advancements evidenced by the Twins (Table 35). These visitors had learned that virtual humans (or advanced AI) exist now and can be created by computer scientists, how a virtual human works, or that technology, in general, has "come a long way." Some answers in this category include:

Some computers can actually understand if you talk to them.
That they can use some sort of software that allows them to understand people's voices.
I didn't know we had this sort of thing yet, thought it was science fiction.
I didn't really know that you can program a computer to understand different types of voices and answer questions.
That there are virtual humans here and that they respond because I didn't think they would.
How smart the computer is now.

Table 35: Responses to "*What did you learn [about technology or computers from integrating with the exhibit?*]"(n=103)

Responses	n	Percent*
Technological capabilities and advancements	65	63%
Science, technology, or computer responses *not* about capabilities or advancements	14	14%
How to use the exhibit	10	10%
Personal information/facts about the Twins	9	9%
Future or current uses for Virtual Humans	3	3%
Positive attitudes towards VH/computers/tech	3	3%
"I don't know"	2	2%
Other/un-codeable	4	4%

* Multiple responses allowed. Percentages total more than 100.

Additional analysis of the data revealed the following trends:

- Those in the Direct condition were much more likely to give a response related to the **technological capabilities and advancements** (75% compared to 54% in the Blended condition; χ^2= 4.793, N=103, p=0.029). Perhaps the unmediated nature of the Direct Interaction made the capabilities and abilities of the Twins stand out more clearly to visitors.
- Participants who gave a more general response about **science, technology or computers**, were more likely to have a higher self-reported knowledge of computers (Mann-Whitney U=411, N=103, p=0.028) and interest in learning about technology (Mann-Whitney U=430.5, N=103, p=0.043) than participants who did not give this response.

STEM Learning Scale of Domains Supported by the Twins

Using the responses to the questions above about the purpose of the exhibit and what, if anything, participants learned at the exhibit about computers or technology, a new variable was created. Participants were scored as 1) having gained *no additional understanding* about STEM domains supported by the Twins or 2) having gained *at least one additional understanding* about STEM domains supported by the Twins.

Analysis of this computed variable revealed that 84% of participants gained at least one additional understanding. Using the independent variables for additional inferential analysis only one significant difference was discovered: adults were significantly more likely than children to have **gained at least one additional understanding** (χ^2= 8.623, N=107, p=0.003); in fact all of those who gained no additional understanding of STEM domains related to the Twins were children. This indicates that the exhibit was more successful at supporting knowledge gain among adults than children. At the same time, however, adults have great life experience and cognitive abilities, and may be better positioned to recognize and articulate "new" learning when it occurs.

Follow-up Questionnaire Results Relative to Learning about Technology and Computers at the Exhibit

Participants in the follow-up questionnaire were asked what they learned about technology or computers from interacting with the Twins. The follow-up responses were similar to the onsite responses, with the category of the technological capabilities and advancements evidenced by the Twins as the most common response (Table 36).

Table 36: Responses to "*What did you learn [about technology or computers from interacting with the exhibit?*]"(n=49)

Responses	n	Percent*
Technological capabilities and advancements	35	71%
How to use the exhibit	4	8%
Positive attitudes towards VH/computers/tech	4	8%
Personal information/facts about the Twins	3	6%
Future or current uses for Virtual Humans	3	6%
Science, technology, or computer responses *not* about capabilities or advancements	2	4%
"I don't know"	1	2%
Other/un-codeable	1	2%
Participant indicated they learned nothing new	3	6%

* Multiple responses allowed. Percentages total more than 100.

Those in the Direct condition were much more likely to give a response related to the **technological capabilities and advancements** on the follow-up questionnaire (84% compared to 58% in the Blended condition; χ^2= 3.953, N=49, p=0.047). This mirrors the findings from the onsite interview.

Findings: Coach Mike and Robot Park

Within the following findings sections focused on Coach Mike and Robot Park, first a description of the sample is given, followed by the visitor impact areas (Engagement and Interest, Attitude, Awareness, and Knowledge) organized by the individual indicators.

Description of Sample

Onsite Observations and Interviews Demographics

A total of 269 visitors were observed at Robot Park. Of these visitors, 46 did not complete the interview, resulting in 223 observations matched with interviews. As summarized in Table 37, about half of the sample was male and half was female. The majority of participants (68%) were under 16 years of age; the mean age of participants categorized as "children" was 10.2 years old. Adults, over the age of 16, represented 32% of all participants, with a mean age of 30.3 years. Visitor groups primarily contained children and on average included 6 people. A little over three fourths of the visitors live in the Northeast (79%), with exactly half residing in Massachusetts (50%).

The treatment group is comprised of those participants who used Robot Park with Coach Mike, and the control group used Robot Park without Coach Mike engaged. When comparing treatment and control groups, there were significantly more adults attending Robot Park without Coach Mike engaged (χ^2= 7.622, N=268, p=0.006). There were no statistically significant differences between the treatment and control groups in terms of gender, geographical region, country, group type, group size, or familiarity with a science field.

Table 37: Onsite Observations and Interviews Sample Description

Sample Description	Overall (n=269)	Conditions	
		Treatment (n=145)	Control (n=124)
Gender of Participant			
Male	58%	64%	52%
Female	42%	36%	48%
Live in Massachusetts?	50%	54%	45%
Live in the Northeast?	79%	77%	80%
Live outside of US?	9%	8%	10%
Age of Participant			
Adult (16 and older)	31%	24%	40%
Child (under 16)	69%	76%	60%
Group Type			
Adult only group	13%	10%	17%
Children in group	87%	90%	84%
Mean Group Size	6 people	5.5 people	6.8 people

Follow-up Online Questionnaire Demographics

A total of 75 visitors responded to the follow-up questionnaire. As shown in Table 38, almost two-thirds of the sample was male and one-third was female. The majority of participants (64%) were under 16 years of age; adults over the age of 16 represented 36% of all participants. Visitor groups primarily contained children and on average included 5 people. Over three fourths of the visitors live in the Northeast (81%), with about half residing in Massachusetts (52%). Overall, the follow-up sample was representative of the onsite sample.

Table 38: Follow-up Online Questionnaire Sample Description

Sample Description	Overall (n=75)	Conditions	
		Treatment (n=35)	Control (n=40)
Gender of Participant			
Male	61%	60%	62%
Female	39%	40%	38%
Live in Massachusetts?	52%	51%	45%
Live in the Northeast?	81%	77%	84%
Live outside of US?	7%	9%	5%
Age of Participant			
Adult (16 and older)	36%	29%	43%
Child (under 16)	64%	71%	57%
Group Type			
Adult only group	12%	10%	17%
Children in group	88%	90%	83%
Mean Group Size	5.3 people	4.6 people	6 people

Coach Mike and Robot Park: Visitor Engagement and Interest

Table 39 shows the Engagement and Interest impacts and indicators for Coach Mike and the Robot Park exhibit. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 39: Coach Mike Engagement and Interest High-Level Results

Impact (Shaded) and its Related Indicators	Level of Evidence Based on the Summative Evaluation
Children (ages 7 -14) and adults will increase their engagement and interest in computer science and technology.	*Somewhat achieved: 2 of 4 indicators showed evidence of increased engagement and interest.*
Visitors to Robot Park will **engage more deeply** when Coach Mike is turned on, with the following indicators of engagement: 1) stay time at the exhibit, 2) number of programs created, 3) length of programs created, 4) completion of a task/goal incorporated into the exhibit.	• There was a significant statistical difference in the stay times between the treatment and control conditions, with the treatment group staying at the exhibit longer. • There was no significant statistical difference between number of programs created or program length of those who interacted with the exhibit when Coach Mike was engaged (treatment) and those who did not have Coach Mike engaged (control).

	• There was a significant statistical difference in the completion of the sign-related goal between the treatment and control conditions, with the treatment group completing the goal more often.
Visitors will indicate that they had a **positive experience** at the exhibit.	• There were no significant differences between treatment and control groups in how they rated their exhibit experience
Visitors will indicate their **interest in learning more** about computer science.	• There were no significant differences between treatment and control groups in how they rated their interest in learning more about computers by interacting with the exhibit.
Visitors will have a **conversation after they leave** the museum about the exhibit experience.	• 57% of respondents to the follow-up questionnaire indicated that they did talk to someone after leaving the museum, with family members the most common people to have a conversation with.

Visitors to Robot Park will engage more deeply when Coach Mike is turned on, with the following indicators of engagement: 1) stay time at the exhibit, 2) number of programs created, 3) length of programs created, 4) completion of a task/goal incorporated into the exhibit (Engagement and Interest Indicator)

Stay time at the Robot Park exhibit

Time spent at the exhibit ranged from 9 seconds to just over 25 minutes, with a mean time of 4 minutes and 27 seconds (Table 40). When looking at stay time between the treatment and control groups, there was a significant statistical difference between conditions (independent samples t-test; t=2.003, N=269, p=.046), with visitors in the treatment group spending more time at the exhibit on average.[6] There were no significant differences in stay time based on visitor age or sex.[7]

[6] A parametric test (i.e. a t-test) was used for this analysis, although the data were not collected using a random violating one of the assumptions for a t-test.

[7] No significant differences were found using t-tests. Similarly, an ANCOVA that incorporated condition (treatment/control), sex (male/female), and age (adult/child) as variables yielded no significant differences between groups.

Table 40: Stay Time at Robot Park

Stay Time	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	269	4 min 27 sec	3 min 32 sec	3 min 37 sec	9 sec	25 min 14 sec
Treatment	145	4 min 51 sec	3 min 44 sec	4 min 12 sec	29 sec	25 min 14 sec
Control	124	4 min	3 min 17 sec	2 min 44 sec	9 sec	13 min 50 sec

In an effort to understand stay time in more depth, a cluster analysis was undertaken to see if visitors' stay time clustered naturally into groups. The results indicated two clusters: Group 1 (n=29) clustered around 7 minutes and 41 seconds; Group 2 (n=240) clustered around 3 minutes and 28 seconds. Additional analysis was undertaken around these two clusters, as reported below:

- **Visitor Demographics of Group 1**: Visitors in this cluster generally resembled the larger study sample. Of the 29 visitors in this cluster, 5 were adults and 24 were children (with a cell size too small to test for significant difference within Condition); 10 were female and 19 were male with no significant differences between control and treatment (p=.392).
- **Stay Time within Condition**: Within Group 2, there was no statistically significant difference in stay time between treatment and control (using an independent samples t-test). Within Group 1 (those with a longer overall stay time), there was a statistically significant difference in stay time between treatment and control (independent samples t-test; t=2.169, N=29, p=.039). Within Group 1, treatment visitors stay significantly longer than control visitors (a mean of 13 minutes and 38 seconds for treatment compared to mean of 10 minutes and 52 seconds).
- **Social Interactions**: Visitors in Group 1 were significantly more likely to have a social interaction with another visitor than were visitors in Group 2 (χ^2= 5.022, N=269, p=.025*)*. Group 1 visitors also were significantly more likely to have a social interaction with a staff member than were visitors in Group 2 (χ^2= 8.784, N=269, p=.003).
- **Completion of Challenges**: Visitors in Group 1 were significantly more likely to complete the light up the sign challenge (χ^2= 31.585, N=269, p<.000), the make a square challenge (χ^2= 55.778, N=145, p<.000), and the turn-around challenge (χ^2= 63.283, N=145, p<.000) than were visitors in Group 2. Within Group 1, 25 completed the light up the sign challenge (with a cell size too small to test for significant difference within Condition). When considering the challenges that only treatment visitors could attempt, 9 out of the 19 treatment visitors in Group 1 completed the make a sign challenge and 10 of 19 completed the turn-around challenge. The majority of those visitors completing challenges, therefore, were drawn from this cluster who spent more time at the exhibit (Refer to Table 43 for the main findings related to completing challenges).

Number of programs created

At Robot Park, visitors create computer programs using TERN, a tangible programming language. For this evaluation, visitors created programs and then executed them by pressing the "Run" button. Observations included unintentional pushes of the "Run" button, pushing "Run" for a program the visitor who is being observed did not create, and when other visitors pushed "Run" while the visitor being observed is actively programming the robot.

As shown in Table 41, participants created between zero and 47 programs. The median number of programs run was three. When number of programs created was compared by treatment and control,

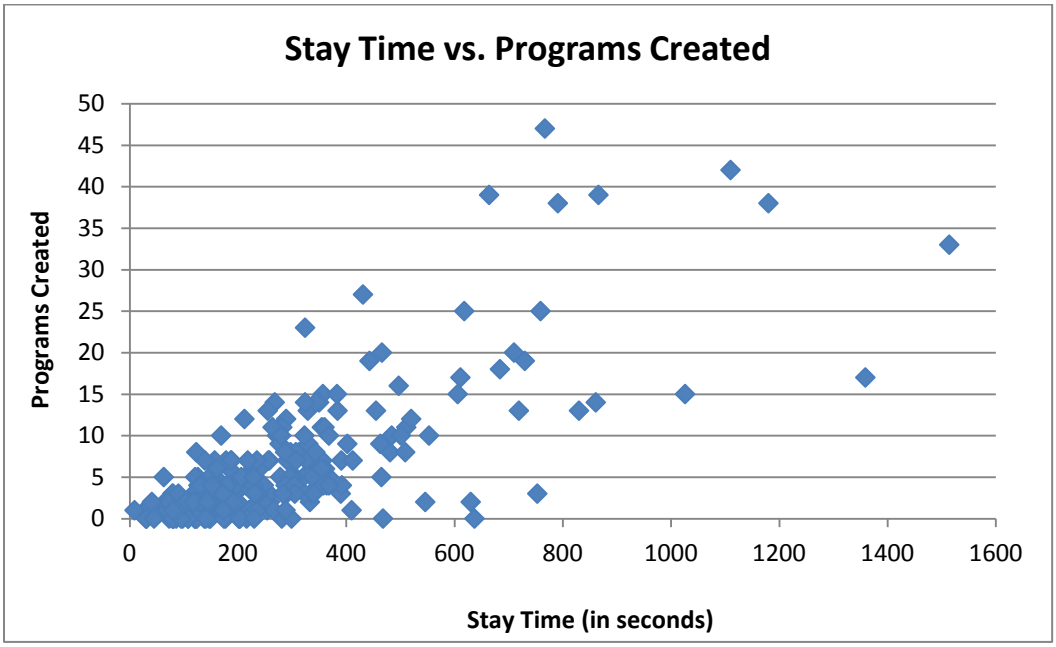
then age group (adults and children), then gender, no significant differences were discovered (Mann-Whitney U test). [8]

Table 41: Number of Programs Created

# Programs	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	269	5.7	3	7.5	0	47
Treatment	145	6.2	3	8	0	42
Control	124	5.3	3	6.9	0	47

A significant relationship emerged when comparing time spent and number of programs created. There is a strong positive correlation between time spent at the exhibit and the number of programs visitors run (Spearman's correlation= 0.704, N=269, *p*=0.000). As stay time increased, the number of programs visitors created increased (Figure 1).

Figure 1: Scatterplot of Stay Time and Programs Created



Length of programs created

Researchers observed the length of each program created by participants. The program length was a count of blocks that included the start block and any other blocks that appeared on the monitor as an

[8] An ANCOVA that incorporated the number of programs created, sex (male/female), and age (adult/child) as variables yielded no significant differences between groups.

active part of the program. Visitors created programs using up to 17 blocks. The average program length was 6.02 blocks (Table 42).

Table 42: Average Program Length

Program Length	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	268	6.02	5.61	4.07	0	17
Treatment	145	5.95	5.33	3.94	0	17
Control	123	6.09	5.73	4.24	0	16

Researchers looked for significant differences between treatment and control, based on age group (adults and children), and between males and females. One significant difference emerged. Females had the highest program length. There is a statistically significant difference between male and females median program length (Mann-Whitney U=4973, N=232, p=0.003). It can be further concluded that males elicited statistically significant lower program lengths than females.

Program length was compared with stay time at the exhibit to determine whether time spent in the exhibition was related to program length. There is a weak negative relationship between time spent at the exhibit and the average length of programs visitors run (Spearman's correlation= -0.174, N=236, p=0.007). When time spent at the exhibit increased, the average length of programs decreased; visitors appear to be spending their time revising and creating new programs as opposed to concentrating on the length of programs.

Completion of a task/goal incorporated into the exhibit

At Robot Park, visitors may attempt up to three challenges: 1) light up the sign, 2) turn around, and 3) make a square. When activated, Coach Mike challenges visitors to solve these specific problems. A visitor successfully completes the sign challenge when they move the robot to the "Target" sign and the overhead "Robot Park" sign lights up and moves. Turning around is a task issued by Coach Mike. It involves moving the robot in a tight circle without the use of the "spin" block. Make a square is another challenge issued by Coach Mike and involves moving the robot in a square. Researchers recorded when visitors attempted and/or completed these challenges.

All visitors to Robot Park had the opportunity to light up the sign. About half (44%) of all visitors attempted to light up the Robot Park sign; Half (50%) of the treatment group and just over one-third (36%) of the control group attempted this challenge. Only eight percent of visitors in the treatment group attempted the challenges issued by Coach Mike[9]. The completion rate for challenges was quite high. Almost all visitors who attempted a challenge managed to complete it (See Table 43).

[9] All those who attempted either the "turn around" or "make a square" challenge used the Coach Mike help button; a press of the button was needed to issue the challenge. However, of the 97 individuals who used the Coach Mike button, the majority (88%) did not attempt a challenge.

Table 43: Completed Challenges at Robot Park

Completed Challenges	Overall Exhibit			Treatment			Control		
	Attempt.	Comp.	%	Attempt.	Comp.	%	Attempt.	Comp.	%
Light up sign	117	103	88%	72	64	88%	45	39	87%
Turn around	11	11	100%	11	11	100%	n/a	n/a	n/a
Make a square	11	10	91%	11	10	91%	n/a	n/a	n/a

Note. Visitors could attempt and/or complete more than one challenge. A total of 12 individuals attempted the "turn around" or "make a square" challenges; 10 of these 12 tried both and 2 individuals attempted one or the other of the challenges.

The likelihood that a visitor would attempt to light up the sign was dependent on the condition (χ^2= 4.858, N=269, p=0.028); treatment visitors were more likely to attempt the sign challenge, see Table 44. Also, treatment visitors were more likely to complete the task of lighting up the Robot Park sign (χ^2= 4.553, N=269, p=0.033). A 95% confidence interval shows that between 1% and 24% more visitors will complete the sign challenge if Coach Mike is engaged. For the turn around and square challenge, no comparisons between treatment and control were conducted as these two challenges were exclusive to the treatment group and required Coach Mike to be engaged.

Table 44: Light up Sign by Condition

Challenge	Treatment (n=145)		Control (n=124)		χ^2 value	p value
	n	Percent	n	Percent		
Light up sign: attempted	72	50%	45	36%	4.858	.028*
Light up sign: completed	64	44%	39	31%	4.553	.033*

Note. The Chi Square test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

When comparing challenges to additional variables such as age, gender, number of programs created, and program length, there were several significant findings. Males in the control condition were significantly more likely to complete the light up the sign challenge than were females in the control condiditon (χ^2= 7.104, N=122, p=0.008). As shown in Table 45, visitors attempting challenges were more likely to run more programs than visitors no attempting challenges (Mann-Whitney U=4185.5, N=269, p=0.000).

Table 45: Challenge Attempts Compared with Number of Programs Created

Challenges attempted	Number of Programs: Mean Rank		Sample Size		*U* value	*p* value
	Attempted	Did not attempt	Attempted	Did not attempt		
Attempted at least one challenge	175.06	103.69	118	151	4181.5	.000**
Light up sign	174.66	104.47	117	152	4252	.000**
Turn around	132.14	68.15	11	134	86.5	.000**
Make a square	131.82	68.17	11	134	90	.000**

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Those visitors that attempted to turn around the robot or make a square spent a significantly longer amount of time at the exhibit than those visitors not attempting these challenges (Mann-Whitney *U*=88, N=145, *p*=0.000; Mann-Whitney *U*=55, N=145, *p*=0.000). Visitors who attempted the task of turning around the robot or the sign challenge were also more likely to run shorter programs; however, there was no significant difference in program length for visitors who attempted to make a square (Table 46). This indicated that visitors attempting to solve challenges were focused and created multiple, but compact, programs in an attempt to achieve their goal. They focused on multiple iterations as opposed to concentrating on the length of programs. This finding draws into question the hypothesis that more or longer programs would be the result of the treatment condition. The team may want to form and test new hypotheses for what success looks like with Coach Mike. It may be that a measure of success for Coach Mike participants would be the creation of multiple, shorter programs.

Table 46: Challenge Attempts Compared with Program Length

Challenges attempted	Program Length: Mean Rank		Sample Size		*U* value	*p* value
	Attempted	Did not attempt	Attempted	Did not attempt		
Attempted at least one challenge	101.48	134.13	113	123	5026	.000**
Light up sign	102.23	133.19	112	124	5122	.001**
Turn around	44.36	68.51	11	121	422	.045*
Make a square	49.73	68.02	11	121	481	.129

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Observed Behaviors and Usability Issues

As visitors moved through the exhibit, data collectors recorded the occurrence of selected behaviors:

- *Looks at the screen*: the target visitor looks at the monitor screen mounted on the back wall of Robot Park.
- *Physically moves the robot*: visitor picks up and moves the robot during the observation
- *Uses block tester*: whenever the visitor places at least one block in the "Block Tester" area of the exhibit
- *Other*: any other generally observable behavior that the data collector feels is important

As illustrated by Table 47, the most frequent behavior was "looks at screen," exhibited by 88% of all sampled visitors. This behavior applies when Coach Mike is off and also when Coach Mike is engaged and talking. The next most frequent behavior was using the "block tester," which was exhibited by 62% of the visitors. This was followed by "moves robot," exhibited by 11% of the visitors. All the other behaviors occurred less frequently.

Table 47: Observed Behaviors at Robot Park

Behaviors	Overall Exhibit (n=269)		Treatment (n=145)		Control (n=124)	
	n	Percent	n	Percent	n	Percent
Looks at screen	236	88%	124	86%	112	90%
Block tester	166	62%	93	64%	73	59%
Moves robot	29	11%	16	11%	13	11%
Other	17	6%	2	1%	15	12%

When examining the relationship between observed behaviors and exhibit condition, one significant finding emerged. Visitors who attended Robot Park without Coach Mike engaged were more likely to demonstrate "other" behaviors, including using the block tester for the majority of movements (as opposed to creating a program), and pushing run without the start block or without creating a program (χ^2= 12.968, N=269, p=0.000). These specific behaviors reflect visitors' misunderstanding of how to execute the TERN programming language. While engaged, Coach Mike provides tips on how to start and successfully complete a program.

A relationship also exists between observed behaviors and stay time. Visitors who looked at the screen, used the block tester, and moved the robot spent more time at the exhibit (Table 48). As expected, visitors who engaged with the exhibit more deeply demonstrated longer stay times. Studies have shown a positive correlation among time and visitor behaviors (Serrell, 1998).

Table 48: Observed Behaviors Compared with Stay Time (n=269)

Behaviors	Stay Time: Mean Rank		Sample Size		*U* value	*p* value
	Behavior Observed	Behavior not observed	Behavior Observed	Behavior not observed		
Looks at screen	140.20	97.79	236	33	2666	.003**
Block tester	14506	118.79	166	103	6879	.007**
Moves robot	187.07	128.71	29	240	1970	.000**
Other	152.94	133.79	17	252	1837	.326

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

For observed behaviors, a comparison between gender and between adults and children showed no significant differences.

Data collectors recorded usability issues, including hardware and user errors, which occurred while the visitor being observed was at Robot Park. The most frequent issue users faced was the mat. The mat functioned as an on-off sensor for Coach Mike – when a visitor stepped on the mat, Coach Mike would activate and greet the visitor, and as the visitor stepped off the mat, he would say "goodbye." This feature caused usability issues for visitors. Almost one quarter of visitors (24%) were not fully on the mat, did not weigh enough to trigger the mat, or stepped on and off the mat repeatedly while still at the exhibit. This behavior would cause Coach Mike to start over, preventing him from fully functioning as an intelligent tutor. Another usability issue visitors experienced was with the camera, which was integral to sending the program to the robot. Some visitors blocked the camera (20%), while others encountered technical issues where the camera was unable to read blocks that were worn or unreadable due to glare (16%) (Table 49). A small percentage of visitors experienced a more substantial usability issue where the computer crashed (3%). Other usability issues included problems with the block tester and trouble with the Coach Mike button responding.

Table 49: Usability Issues (n=269)

Usability Issues	n	Percent
Mat*	65	24%
Camera	55	20%
Problem reading blocks	44	16%
Frozen computer/crash	7	3%
Other	28	10%

* The mat was used for the treatment or Coach Mike condition only; stepping on the mat triggered Coach Mike to welcome the visitor and begin his tutorial.

When comparing usability issues between treatment and control groups, visitors attending Robot Park with Coach Mike engaged were more likely to have issues with the camera (Table 50). There was also a

significant difference between treatment and control for "other" usability issues. As many of these "other" issues related to the Coach Mike button, it is logical that the treatment group would have these problems more frequently.

Table 50: Usability Issues by Condition

Interactions	Treatment (n=145)		Control (n=124)		χ^2 value	*p* value
	n	Percent in group	n	Percent in group		
Mat	65	45%	-	-	-	-
Camera	41	28%	14	11%	11.856	.001**
Problem reading blocks	23	16%	21	17%	0.056	.812
Frozen computer/crash	1	1%	6	5%	4.540	.051
Other	24	17%	4	3%	12.728	.000**

Note. The Pearson Chi-Square test was used to test for statistical significance. Fisher's exact test was also applied and yielded the exact same results.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Group type also contributed to usability issues. When children were present in a visiting group, the target visitor was more likely to have issues with the mat (χ^2= 5.102, N=263, *p*=0.024). Statistical tests to compare usability issues across gender and age category yielded no significant differences.

Treatment Visitors' Interactions with Coach Mike

Researchers observed treatment visitors' behaviors related to Coach Mike specifically. These behaviors were used as indicators that visitors noticed and paid attention to Coach Mike. Almost all of the treatment visitors (90%) interacted with Coach Mike in some way, as indicated by the "overall" variable (Table 51). The most frequent behavior observed was a visual reference to the screen. Eighty-six percent of visitors interacted with Coach Mike by looking up at the screen when he was talking. A high percentage of visitors also were observed pushing the Coach Mike help/hint button; 67% of visitors exhibited this behavior. An equal percentage of visitors (68%) also attempted to follow Coach Mike's directions. Other interactions with Coach Mike occurred less frequently: 19% verbally communicated with Mike (i.e. directed comments at Mike), 16% mentioned Mike to another visitor, and 3% exhibited other behaviors.

Table 51: Coach Mike Interactions

Interactions	n	Percent*
Overall	131	90%
Looks at screen	124	86%
Mike button	97	67%
Tries to follow Mike's directions	97	67%
Verbally communicates with Mike	27	19%
Mentions to another visitor	23	16%
Other	5	3%

*Multiple interactions allowed. Percents may add up to more than 100.

Researchers tested whether age affected treatment visitors' interactions with Coach Mike; several significant findings emerged (Table 52). Adults were more likely than children to mention Coach Mike to another visitor or staff member (χ^2= 11.734, N=145, p = 0.001). Adults also verbally communicated with Coach Mike more frequently than did children (χ^2=4.994, N=145, p = 0.025). This includes verbally answering Coach Mike, talking to him, and repeating his directions out loud.

Table 52: Mike Interactions by Age Group

Interactions	Adult (n=35)		Child (n=110)		χ^2 value	p value
	n	Percent in group	n	Percent in group		
Overall	33	94%	98	89%	0.821	.365
Looks at screen	33	94%	91	83%	2.864	.091
Mike button	25	71%	72	65%	0.428	.513
Follow Mike's directions	26	74%	71	65%	1.138	.286
Communicates with Mike	11	31%	16	15%	4.994	.025*
Mentions to another visitor	12	34%	11	10%	11.734	.001**
Other	2	6%	3	3%	0.712	.399

Note. The Pearson Chi-Square test was used to test for statistical significance. Fisher's exact test was also applied and yielded the exact same results.
* Significant at the p<.05 level. **Significant at the p<.01 level.

There was a statistically significant difference between the appearance of certain behaviors and the median stay time in the exhibit. Visitors who do not look at the screen when Coach Mike is engaged spend a statistically significant lower amount of time at the exhibit than visitors who look at the screen (Mann-Whitney U=850, N=145, p=0.011). Visitors who pressed the Coach Mike help button, followed Mike's directions, and communicated with Mike also demonstrated higher stay times (Table 53).

Table 53: Mike Interactions Compared with Stay Time (n=145)

Interactions	Stay Time: Mean Rank		Sample Size		*U* value	*p* value
	Behavior Observed	Behavior not observed	Behavior Observed	Behavior not observed		
Overall	75.03	53.96	131	14	650.5	.074
Looks at screen	76.65	51.48	124	21	850	.011*
Mike button	79.36	60.15	97	48	1711	.010*
Follow Mike's directions	81.04	56.76	97	48	1548.5	.001**
Communicates with Mike	94.80	68.01	27	118	1004.5	.003**
Mentions to another visitor	87.13	70.34	23	122	1078	.079
Other	83.00	72.64	5	140	300	.588

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Interactions with Coach Mike as a whole (i.e. the "overall" variable) did not affect the number of programs created; however, visitors who used the Coach Mike help button did create longer programs (Mann-Whitney *U*=1481, N=132, *p*=0.036).

Visitors pressed the Mike button on average about two times. The number of visitor pushes ranged from one to seven. There were no differences between adults and children regarding their usage of the help button. Visitors who set goals with other visitors pushed the Coach Mike help button more frequently (Mann-Whitney *U*=889, N=97, *p*=0.041). There was also a weak positive relationship between the number of programs created and number of times visitors pressed the Coach Mike help/hint button (Spearman's correlation= 0.276, N=269, *p*=0.006). The more frequently a visitor used the Coach Mike help button, the more programs he/she created. There was a small negative relationship between the average length of programs and number of times visitors pressed the help button (Spearman's correlation= -0.259, N=236, *p*=0.014); as the usage of the Mike button increased, the average length of the programs decreased. One way to interpret this finding is that visitors who specifically sought assistance from Coach Mike spent more of their time revising and creating new programs, as opposed to concentrating on the length of programs.

Visitors' Interactions with Other Visitors and Staff

Social Interactions between Study Participants and Other Visitors

One measure of social interaction is the way in which visitors interact with each other. Data collectors observed interactions between the target visitor and any other visitor, whether they were in the target's group or not. Six categories of interactions were defined:

- *Conceptual*: target visitor talks with visitors about the purpose of the exhibit or concepts the exhibit is trying to teach.
- *Operational*: target visitor talks with visitors about how to operate the exhibit.
- *Goal Setting*: target visitor talks with visitors about completing the challenge or setting another goal.
- *Positive Affect*: any positive talk or utterance.
- *Negative Affect*: any negative talk or utterance.
- *Other*: any interaction with visitors that does not clearly fit into one of the previous categories.

These categories were used to create an "overall" variable which indicated any social interaction with another visitor, regardless of type.

The majority of visitors interacted with another visitor at the exhibit (87%)[10]. Table 54 displays the frequency for each type of visitor interaction.

Table 54: Types of Interactions between the Visitor being Observed and another Visitor

Interactions	n	Percent*
Overall	233	87%
Operational	190	71%
Positive Affect	178	66%
Goal Setting	128	48%
Conceptual	77	29%
Negative Affect	66	25%
Other	4	2%

*Multiple interactions allowed. Percents may add up to more than 100.

When looking at all interactions with other visitors (i.e. the "overall" variable), findings included:

- Any interaction with another visitor increased the target visitor's **stay time** at the exhibit; those interacting with another visitor spent an average of 4 minutes, 48 seconds compared to an average of 2 minutes, 15 second for those who did not interact with other visitors (Mann-Whitney U=1985.0, N=269, p=0.000). This increase in stay time was found regardless of condition.
- Related to the above finding, any interaction with another visitor significantly increased the **number of programs** a visitor created (6.3 programs on average compared to 2.1 programs without a staff interactions; Mann-Whitney U=2080.0, N=269, p=0.000).
- **Adult study participants** (i.e. the target of the observation) were significantly more likely than child participants to have any interaction with another visitor (94% compared to 83%; χ^2= 5.888, N=268, p = 0.015).
- Varying the condition type (control versus treatment) did not significantly impact the amount of social interaction overall between visitors.

[10] Formative evaluation conducted at Robot Park by MOS staff found a lower overall rate of interaction between the target visitor and other visitors, with 41% of observed visitors interacting with another visitor; Bronnenkant, K., & Cotterill, S. (2011). *"RobotPark" Front-End Evaluation Results*. Unpublished Memorandum from MOS staff to ICT staff.

Analyses were also performed within each of the distinct interaction types.

- **Operational talk**:
    - Operational talk between visitors was related to an increased the **stay time** of the target visitor (Mann-Whitney *U*=4278.0, N=269, *p*=0.000) and a higher average **number of programs created (**Mann-Whitney *U*=5110.5, N=269, *p*=0.000).
    - Operational talk between the target and another visitor was more likely to occur in the **control/Robot Park only setting** (77% compared to 66%; χ^2= 3.967, N=269, *p* = 0.046). This is likely related to the type of scaffolding Coach Mike provides, which is primarily operational. In the absence of Coach Mike, visitors take up this type of talk.
- **Goal-setting talk** between visitors was related to an increased the **stay time** of the target visitor (Mann-Whitney *U*=4201.0, N=269, *p*=0.000) and a higher average **number of programs created (**Mann-Whitney *U*=3891.0, N=269, *p*=0.000).
- **Conceptual talk** between visitors was related to an increased the **stay time** of the target visitor (Mann-Whitney *U*=5974.0, N=269, *p*=0.014).
- **Positive affective talk** between visitors was related to an increased the **stay time** of the target visitor (Mann-Whitney *U*=4294.0, N=269, *p*=0.000) and a higher average **number of programs created (**Mann-Whitney *U*=4390.0, N=269, *p*=0.000).
- Perhaps counter-intuitively**, negative affective talk** between visitors was related to an increased the **stay time** of the target visitor (Mann-Whitney *U*=4347.5, N=269, *p*=0.000) and a higher average **number of programs created (**Mann-Whitney *U*=5585.0, N=269, *p*=0.041).

Researchers tested relationships between visitor interactions overall and Coach Mike-related behaviors. When the target visitor interacts with other visitors, they are more likely to mention Coach Mike (χ^2= 4. 374, N=145, *p* = 0.036); naturally, visitor interaction provides an opportunity for visitors to share information. Visitor interaction also increases the chances that visitors communicate directly with Coach Mike (χ^2= 5. 308, N=145, *p* = 0.021).

When looking at specific types of visitor interactions, some significant findings emerged. Visitors demonstrating conceptual talk are more likely to mention Mike to another visitor (χ^2= 8.377, N=145, *p* = 0.004). Visitors who talk about completing the challenge or setting another goal are more likely to communicate directly with Mike (χ^2= 12.059, N=145, *p* = 0.001).

As shown in Table 55, visitors who made a positive comment (i.e. “You did it!, “Cool!”, “Yes!”) mentioned Coach Mike to other visitors more frequently (χ^2= 10.593, N=145, *p* = 0.001). They also communicated with Mike more than those who did not exhibit a positive response (χ^2= 13.426, N=145, *p* = 0.000). Similarly, visitors making negative comments or utterances (Table 56) were more likely to interact with Coach Mike through verbal communication (Table 52; χ^2= 13.859, N=145, *p* = 0.000).

Table 55: Mike Interactions by Positive Affect to another Visitor

Interactions	Positive affect (n=96)		No positive affect (n=49)		χ^2 value	*p* value
	n	Percent in group	n	Percent in group		
Overall	88	92%	43	88%	0.569	.451
Looks at screen	84	88%	40	82%	0.902	.342
Mike button	65	68%	32	65%	0.085	.771
Tries to follow Mike's directions	66	69%	31	63%	0.441	.507
Communicates with Mike	26	27%	1	2%	13.426	.000**
Mentions to another visitor	22	23%	1	2%	10.593	.001**

Note. The Pearson Chi-Square test was used to test for statistical significance. Fisher's exact test was also applied and yielded the exact same results.
* Significant at the p<.05 level. **Significant at the p<.01 level.

Table 56: Mike Interactions by Negative Affect to another Visitor

Interactions	Negative affect (n=39)		No negative affect (n=106)		χ^2 value	*p* value
	N	Percent in group	N	Percent in group		
Overall	38	97%	93	88%	3.075	.079
Looks at screen	37	95%	87	82%	3.769	.052
Mike button	28	72%	69	65%	0.578	.447
Tries to follow Mike's directions	29	74%	68	64%	1.342	.247
Communicates with Mike	15	39%	12	11%	13.859	.000**
Mentions to another visitor	10	26%	13	12%	3.823	.051

Note. The Pearson Chi-Square test was used to test for statistical significance. Fisher's exact test was also applied and yielded the exact same results.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Social Interactions between Study Participants and Staff Members

Another measure of social interaction includes verbal interactions between the target visitor being observed and staff members or volunteers at Cahners Computer Place. Six categories of staff interactions were defined:

- *Conceptual*: target visitor talks with a staff member about the purpose of the exhibit or concepts the exhibit is trying to teach.
- *Operational*: target visitor talks with a staff member about how to operate the exhibit.
- *Goal Setting*: target visitor talks with a staff member about completing the challenge or setting another goal.
- *Positive Affect*: any positive talk or utterance.
- *Negative Affect*: any negative talk or utterance.
- *Other*: any interaction with a staff member that does not clearly fit into one of the previous categories.

These categories were used to create an "overall" variable which indicated any social interaction with a staff member, regardless of type.

Very few visitors, only 16%, interacted with a Cahners Computer Place staff member or volunteer while at Robot Park (Table 57)[11]. When interacting with staff members, the majority of visitors and staff (15%) had operational conversations; visitors may have talked about how to operate the exhibit with staff or received step-by-step questions or instructions to help complete a goal.

Table 57: Types of Interactions between the Visitor being Observed and Staff Members/Volunteers

Interactions	n	Percent*
Overall	42	16%
Operational	39	15%
Goal Setting	19	7%
Conceptual	17	6%
Positive Affect	14	5%
Negative Affect	14	5%
Other	5	2%

*Multiple interactions allowed. Percents may add up to more than 100.

When looking at all interactions with staff in general (i.e. the "overall" variable), some trends emerged:

- As might be expected, interaction with a staff member or volunteer increased **stay time** at the exhibit; those interacting with staff spent an average of 6 minutes 54 seconds compared to an average of 4 minutes for those who did not interact with staff (Mann-Whitney U=2625.5, N=269, p=0.000).
- Interactions with a staff member/volunteer also significantly increased the **number of programs** a visitor created (10.5 programs on average compared to 4.9 programs without a staff interactions; Mann-Whitney U=2772.0, N=269, p=0.000).

[11] Formative evaluation conducted at Robot Park by MOS staff found a much higher overall rate of interaction between staff and visitors, with 58% of observed visitors interacting with a staff member (Bronnenkant & Cotterill, 2011).

- Visitors who interacted with a staff member/volunteer were also more likely to successfully **light up the sign**; 23% of those interacting with a staff member lit up the sign compared to only 11% of those visitors who did not interact with staff (χ^2= 7.487, N=269, p = 0.006).

Researchers looked at social interaction with staff for differences between 1) control and treatment groups, 2) adults and children, 3) the distinct type of interactions, and 4) as an indicator for various interactions with Coach Mike; no significant results emerged.

Visitors will indicate that they had a positive experience at the exhibit (Engagement and Interest Indicator)

In order to learn more from visitors about their experience at Robot Park, visitors were asked a series of open-ended and close-ended questions. The results of these analyses are reported below.

Participants' Ratings for Interacting with the Exhibit

Visitors were asked to rate their experience interacting with the exhibit. Values were assigned to each rating: Boring =1; Just Okay=2; Pretty Good=3; Exciting=4. Overall visitors had a "pretty good" experience with Robot Park with a mean rating of 3.2 (see Table 58). As shown in Table 59, the majority of visitors felt their interaction with the exhibit was "pretty good" or "excellent." Only one visitor (1%) found interacting with the exhibit to be "boring." There were no significant differences between treatment and control groups in how they rated their exhibit experience (Mann-Whitney U=6009.5, N=221, p=0.743).

Table 58: Interaction Rating

Interacting with the exhibit was…	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	222	3.2	3	0.69	1	4
Treatment	115	3.18	3	0.708	2	4
Control	107	3.21	3	0.673	1	4

Table 59: Interaction Rating Frequencies

Interacting with the exhibit was…	Overall (n=222)	Conditions	
		Treatment (n=115)	Control (n=107)
Boring	1%	0%	1%
Just Okay	14%	17%	11%
Pretty Good	50%	47%	53%
Exciting	45%	36%	35%

Some significant relationships emerged when comparing ratings with social interactions. Visitors who interacted with other visitors were more likely to report a positive experience with Robot Park (see Table 60). Specifically, visitors who spoke with other visitors about operational aspects of the exhibit, goal setting, or just general positive feedback about the exhibit rated their interaction with the exhibit higher than did those who did not engage in those types of conversations. Interactions with staff had no effect on visitors' ratings (Mann-Whitney U test).

Table 60: Visitor Interactions Compared with Interaction Rating (n=222)

Interactions with visitors	Rating Mean Rank		Sample Size		*U* value	*p* value
	Behavior Observed	Behavior not observed	Behavior Observed	Behavior not observed		
Overall	114.59	83.43	200	22	1582.5	.018*
Operational	111.58	111.27	165	57	4689	.973
Positive Affect	121.71	87.88	155	67	3610	.000**
Goal Setting	120.93	101.18	116	106	5054.5	.012*
Conceptual	117.29	109.10	65	157	4726	.342
Negative Affect	123.44	107.57	55	167	3936	.081
Other	183.5	111.17	1	221	38.5	.351

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

There is a weak positive correlation between time spent at the exhibit and interaction rating (Spearman's correlation= 0.242, N=222, p=0.000). As stay time increased, visitors rating regarding interacting with the exhibit increased. There is also a relationship between the total number of programs executed and interaction rating (Ordinal Logistic Regression, p=0.008). As the number of programs increased, the level of interest for exhibit interaction also increased. There was no statistical significance between average length of programs and interaction rating.

Researchers also looked at visitors' knowledge and interest in technology and computers as an indicator for the interaction rating. A prior knowledge of technology did not influence visitors experience interacting with the exhibit (Mann-Whitney *U*=5765, N=221, *p*=0.633). This indicates that even visitors with little knowledge of technology still had a "pretty good" experience at Robot Park. The same held true for knowledge of computers (Mann-Whitney *U*=5244, N=221, *p*=0.633), where knowledge was not related to exhibit enjoyment. However, interest ratings in technology and computers did contribute to visitor interaction at Robot Park. As shown in Table 61, a greater interest in both learning about technology and figuring out how computers work increased the rating for exhibit interaction. Visitors who came to the museum with an existing interest in technology and computers were more likely to rate themselves as having a good time at Robot Park.

Table 61: Interest Ratings compared with Interaction Rating (n=221)

Interest Ratings	Rating Mean Rank		Sample Size		*U* value	*p* value
	Boring & Just okay	Pretty good & Exciting	Boring & Just okay	Pretty good & Exciting		
Learning about technology	86.71	115.15	36	184	2455.5	.007**
Figuring out how computers work	85.50	116.75	44	176	2772	.001**

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Participants' Open-ended Description of Interacting with the Exhibit

Visitors were asked what they did at Robot Park to both get them thinking about their experience and to provide insight into how they perceived their interaction at the exhibit. Over half of the visitors either said they put blocks together (59%) or made the robot move (57%). Some examples of these comments include:

I built blocks and I just pressed the run button.
I tried to get the robot to move around with the puzzle pieces.
I commanded it to go forward, whistle, sing, some other stuff.

A smaller portion of visitors mentioned attempting or completing a task (23%). Additional responses included the idea of "programming," a mention of "play," or interacting with Coach Mike or staff (Table 62 summarizes these responses). When comparing what visitors did at the exhibit by condition, no significant relationships emerged.

Table 62: Responses for "*Can you tell me what you did at the exhibit*?"

Responses	Overall Exhibit (n=222)		Treatment (n=115)		Control (n=124)	
	n	Percent*	N	Percent*	n	Percent*
Put blocks together	131	59%	67	58%	64	59%
Made robot move	126	57%	64	56%	62	57%
Attempt task	52	23%	30	26%	22	20%
"Program"	20	9%	10	9%	10	9%
"Play" or general mention of robot	18	8%	8	7%	10	9%
Interact with Mike or staff	9	4%	8	7%	1	1%
Other	28	13%	14	12%	14	13%

* Multiple responses allowed. Percents may add up to more than 100.

Researches also looked at visitors' responses by age group (Table 63). When asked what they did at the exhibit, children responded more frequently with comments about putting blocks or puzzle pieces together (χ^2= 6.802, N=223, p = 0.009). Adults were more likely to use the word "program" and make a connection that actions at Robot Park simulate computer programming. These results are not surprising as "put blocks together" represents a fairly basic response and using a form of the word "program" could be considered a fairly sophisticated response. Thus, we would expect to see more adults describe their interaction by using more sophisticated language.

Table 63: What Visitors did at the Exhibit by Age

Categories	Adult (n=75)		Child (n=148)		χ^2 value	p value
	n	Percent in group	n	Percent in group		
Put blocks together	35	47%	96	65%	6.802	.009**
Made robot move	38	60%	88	51%	1.566	.211
Attempt task	12	16%	40	27%	3.385	.066
"Program"	11	15%	9	6%	4.494	.034*
"Play" or general mention of robot	8	11%	10	7%	1.025	.311
Interact with Mike or staff	4	5%	5	3%	0.491	.483
Other	11	15%	17	12%	0.458	.498

Note. The Pearson Chi-Square test was used to test for statistical significance. Fisher's exact test was also applied and yielded the exact same results.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Social interactions influenced visitors' response to the question "Can you tell me what you did at the exhibit?" (see Appendix 6 for areas of statistical significance). Contrary to our hypothesis, talking with other visitors and/or staff about goal setting did not significantly increase the likelihood that visitors mention attempting or completing a task (Chi Square test, p>0.05).

Relationships existed between stay time and responses for what visitors did at the exhibit. Visitors who described their visit by making the robot move were significantly more likely to spend a longer time at Robot Park (Mann-Whitney *U*=4958.5, N=223, *p*=0.016). The same held true for visitors mentioning attempting or completing a task (Mann-Whitney *U*=3228.5, N=223, *p*=0.003). When visitors responded with making the robot move or attempting a task, they also had created more programs (Mann-Whitney *U*=4454, N=223, *p*=0.000; Mann-Whitney *U*=3002.5, N=223, *p*=0.000). Knowledge and interest scales did not factor into visitor responses.

Visitors will indicate their interest in learning more about computer science (Engagement and Interest Indicator)

When asked if they would be interested in "learning more about computers by interacting with the exhibit", the majority of visitors indicated this would be "pretty good." Less than one third of the overall sample found learning more about computers by interacting with the exhibit to be "boring" or "just okay" (see Tables 64 and 65). There are no significant differences between treatment and control groups (Mann-Whitney *U*=5688.5, N=221, *p*=0.303).

Table 64: Learning more about Computers Rating

Learning more about computers by interacting with the exhibit would be…	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	222	2.98	3	0.829	1	4
Treatment	115	3.03	3	0.847	1	4
Control	107	2.93	3	0.809	1	4

Table 65: Learning more about Computers Frequencies

Rating	Overall (n=222)	Conditions	
		Treatment (n=115)	Control (n=107)
Boring	3%	3%	3%
Just Okay	27%	26%	28%
Pretty Good	40%	37%	43%
Exciting	31%	35%	26%

Researchers tested interest in learning about computer science by age, gender, group type (children present or no children in visitation group), social interactions, stay time, number of programs, and average program length to explore differences by these factors. No statistical significant relationships emerged.

Although previous knowledge of technology and computers did not affect visitors' interest in learning more about computers (Mann-Whitney U tests), interest levels did factor into how visitors rated learning about computers. Visitors with a higher interest in technology (those that rated their interest as "pretty good" or "exciting") rated their interest in learning more about computers at the exhibit significantly higher than those visitors with little interest in technology (those that rated their interest as "boring" or "just okay"; Mann-Whitney *U*=2026, N=220, *p*=0.000). Similarly, respondents who had a higher interest in computers were significantly more likely to have higher interest ratings for "learning more about computers by interacting with the exhibit" (Mann-Whitney *U*=2027.5, N=220, *p*=0.000).

Visitors will have a conversation after they leave the museum about the exhibit experience (Engagement and Interest Indicator)

Follow-up questionnaire participants were asked the closed-ended question: "After you left the museum, did you talk to anyone about the Robot Park exhibit?" Just over half of the respondents (57%) indicated that they did talk to someone after leaving the exhibit. About 40% did not, and 6% did not recall whether they had spoken to someone about the exhibit.[12]

Follow-up questionnaire participants who indicated that they had talked to someone about the exhibit were asked the open-ended question: "Who did you talk to about Robot Park after your visit?" The majority talked with a family member or a person who accompanied them to the museum (Table 66). Others spoke with friends, co-workers, or someone else. Participants were encouraged to identify anyone that they spoke to, thus some mentioned more than one person.

Table 66: Responses to "*Who did you talk to about Robot Park after your visit?"* (n=39)

Responses	n	Percent*
A family member	27	69%
A person who went to the museum with me that day	18	46%
A friend or co-worker	10	26%
Someone else	4	10%

* Multiple responses allowed. Percentages total more than 100.

Respondents who affirmatively responded they talked with someone after the visit were asked the open-ended question, "What about Robot Park did you and the person talk about?" Most (41%) had a positive comment to report, such as, "*We talked about how fun and easy it is to program a robot."* Others described what they did at the exhibit (36%), or discussed the parallel to computer programming (18%): "*How programming the robot was a very nice parallel to real programming that could really open doors for those who were unfamiliar with it"; "We build robots at home, so it was fun to work with one at the museum."*

[12] Due to the cell size, a comparison between control and treatments groups was not possible.

Coach Mike and Robot Park: Visitor Attitudes

Table 67 shows the Attitudes impacts and indicators for Coach Mike and the Robot Park exhibits. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 67: Coach Mike Attitudes High-Level Results

Impact (Shaded) and its Related Indicators	Level of Evidence Based on the Summative Evaluation
Children (ages 7 -14) and adults will have a positive attitude about computer science and technology.	*Somewhat achieved: 2 out of 5 indicators show evidence of positive attitudes.*
Visitors will indicate they **enjoy learning about technology**.	• Participants mean ratings for interest in learning about technology was 3.3 out of 4 and computers 3.2 out of 4.
Visitors will have a **positive attitude towards virtual humans in society**, as indicated by having positive perceptions of: a) interacting with a computer and b) future developments in the field of artificial intelligence.	• Adult participants in the control condition rated the statements for this indicator higher after having interacting with the exhibit. Treatment participants had no change in ratings.
Visitors will indicate they **enjoy programming or working with robots**.	• Participants overall rated their interest in programming the robot at Robot Park a 3 out of 4; there were no differences in the ratings of control and treatment participants.
Visitors to Robot Park will be **less frustrated with the experience** when Coach Mike is turned on.	• There were no significant differences between how the treatment and control groups rated the ease of using the exhibit.
Visitors will describe Coach Mike as either: a) a **positive influence** (supportive, helpful, motivating) or b) recognize his **suggestions as helpful or useful**.	• 66% described Coach Mike in overall "helpful" terms. • 59% of interview participants who interacted with the Coach Mike condition thought he helped them personally in understanding the exhibit. This increased to 75% in the six week follow-up. • Participants described Coach Mike as giving general advice about the exhibit (30%), acting as a teacher, coach, or guide to the exhibit (24%), or using some variation of the word "help" in their description (i.e. helper, helpful, helped) (22%).
Visitors will **enjoy interacting with the Virtual Humans**.	*(Subsumed by the indicator above)*

Respondents were asked to rate their interest in "learning about technology" and "figuring out how computers work." Overall, respondents rated their interest in both areas fairly high (See Table 68). When additional analysis was performed, independent variables such as condition (control/treatment), age (adult/child), time spent, number of programs created, length of programs, and interactions with other visitors or staff yielded no significant differences for either measure.

Table 68: Respondents' Ratings of their Interest in Technology and Computers (*Scale: 1="boring" and 4="exciting"*; n=220)

	How would you describe your interest in:	
	Learning about Technology is…	Figuring out how computers work is…
Mean Rating	3.3	3.2
Boring	3%	3%
Just okay	14%	17%
Pretty good	38%	33%
Exciting	46%	47%

There were no significant differences in respondents' ratings of their interest in computers or technology when the onsite ratings were compared with the online ratings (i.e. post to delayed post). It is not surprising that over the course of six weeks respondents' interest in the topics in general held constant.

Respondents were also asked to rate statements relative to their knowledge of technology and computers. Overall when compared to their interest ratings, respondents rated themselves as more interested than knowledgeable, with mean knowledge ratings below 3.0 (Table 69).

Table 69: Respondents' Ratings of their Knowledge of Technology and Computers (*Scale: 1="nothing" and 4="a lot";* n=221)

	How much would you say you know about:	
	Technology	Computers
Mean Rating	2.7	2.9
Nothing	3%	3%
A little	39%	30%
Quite a bit	39%	41%
A lot!	18%	25%

When additional analysis was performed using the onsite ratings, independent variables such as condition (control/treatment), age (adult/child), and interactions with staff yielded no significant differences on either measure. In fact no significant differences were found regarding knowledge of technology; however, the following areas of significance were found for knowledge of computers.

- There was a strong correlation between **time spent** and knowledge of computers (Spearman's rho=-.214, N=221, *p*=.001), with those answering "a little" and "quite a bit" spending longer on average at the exhibit than respondents on either extreme of the scale.
- There was a strong correlation between **number of programs run** and knowledge of computers (Spearman's rho=-.152, N=221, *p*=.024), with those answering "quite a bit" running more programs on average than other respondents.
- Target visitors who **did not attempt any of the challenges** significantly more likely to rate their knowledge of computers higher (3.3 compared to 2.7 for those who did attempt a challenge; Mann-Whitney *U*=1044.0, N=115, *p*=.000).
- Target visitors who **did not have any social interactions** with other visitors were significantly more likely to rate their knowledge of computers higher (3.3 compared to 2.8 for those who did interact with another visitor; Mann-Whitney *U*=1572.0, N=221, *p*=.021).
- Similarly, target visitors who **did not talk about goal setting** with another visitor were significantly more likely to rate their knowledge of computers higher (3.0 compared to 2.8 for those who did talk about goal setting with another visitor; Mann-Whitney *U*=4859.0, N=221, *p*=.006).

There were no significant differences in respondents' ratings of their knowledge of computers or technology when the onsite ratings were compared with the online ratings (i.e. post to delayed post). It is not surprising that over the course of six weeks respondents' perceived knowledge of the topics held constant.

Visitors will have a positive attitude towards computers/virtual humans in society, as indicated by having positive perceptions of: a) interacting with a computer and b) future developments in the field of artificial intelligence. (Attitude Indicator)

A quantitative approach was used with adults to determine if interacting with the exhibit impacted self-reported agreement with 1) "In the future, there will be new and exciting innovations with smarter computers" and 2) "In the future, interacting with computers will be easier." When looking at adults' ratings overall, regardless of condition, adults had a significantly higher rating of one of the measures of attitudes towards computers/virtual humans after their interaction with the exhibit (See Table 70).

Table 70: Adult Respondents' Ratings (Retrospective-Pre/Post) of their Attitudes towards Computers/Virtual Humans *(Scale: 1="strongly disagree," 7="strongly agree"*; n=73)

Statement	Retrospective-Pre Rating (Mean)	Post Rating (Mean)	*Z value*	*p* value
In the future, interacting with computers will be easier	6.1	6.3	-2.387	.017**
In the future, there will be new and exciting innovations with smarter computers	6.2	6.4	-1.912	.056

Note. The Wilcoxon Signed Ranks test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

When the results were analyzed based on condition, only control participants showed statistically significant changes in their ratings retrospective-pre to post. This indicates that for these measures, Coach Mike *actually hindered* meaning-making regarding attitudes towards computers and virtual humans that the designers intended to support. Robot Park without Coach Mike did support these attitudes. The significance levels for each statement by condition are reported below:

- *Treatment: In the future, interacting with computers will be easier.* No significant change in mean rating retrospective-pre to post (*Z*=-.948, N=28, *p*=.343).
- *Control: In the future, interacting with computers will be easier.* Significant change in mean rating retrospective-pre to post (*Z*=-2.070, N=45, *p*=.038).
- *Treatment: In the future, there will be new and exciting innovations with smarter computers*. No significant change in mean rating retrospective-pre to post (Z=-1.342, N=28, p=.180).
- *Control: In the future, there will be new and exciting innovations with smarter computers.* Significant change in mean rating retrospective-pre to post (*Z*=-1.964, N=45, *p*=.050).

When comparing the onsite ratings with the online questionnaire ratings (i.e. post to delayed post), there were no significant differences in how either of the statements were rated. In other words, respondents' attitudes held constant in the six weeks after the museum visit. This could be interpreted as the gains that were made directly after interacting with the exhibit were maintained overtime. Another possibility is that these results are a "practice effect" from answering the same question multiple times. Also possible is that not enough time had passed for a decay of the effects to have occurred. A questionnaire conducted six months after having used the exhibit may yield different

results. There were no differences between the ratings of the treatment and control groups in how the statements were rated in the delayed post.

Visitors will indicate they like programming or working with robots (Attitudes Indicator)

Visitors to the exhibit were asked to rate their interest in programming the robot. Overall, visitors indicated that they liked programming or working with robots. As shown in Table 71, the median rating for the treatment group was 3, "pretty good," and the median rating for the control group was 4 "exciting." However, there is no significant difference between the median programming rating for visitors in the treatment group, attending Robot Park with Coach Mike engaged, and the control group, Robot Park only (Mann-Whitney U=5321.5, N=221, p=0.078). Table 72 breaks down the frequencies for each rating.

Table 71: Programming Robot Rating

Programming the robot to move was…	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	221	3.24	3	0.843	1	4
Treatment	115	3.16	3	0.844	1	4
Control	107	3.33	4	0.836	1	4

Table 72: Programming Robot Frequencies

Rating	Overall (n=221)	Conditions	
		Treatment (n=115)	Control (n=107)
Boring	5%	4%	5%
Just Okay	13%	16%	9%
Pretty Good	37%	40%	34%
Exciting	46%	40%	52%

Researchers tested whether age, gender, group type, social interactions, stay time, number of programs created, or average program length factored into visitors' programming rating. There are several significant findings. Visitors who talked with other visitors about goal setting or completing a challenge rated their experience programming higher (see Table 73). Also, visitors who exhibited a positive affect were more likely to rate programming the robot higher than those who did not make a positive utterance. Visitors who interacted with staff members rated their programming experience higher than those who did not talk to staff (See Table 74). Specifically, when target visitors talked with staff about how to operate the exhibit, their ratings for programming increased. It seems logical that visitors inquiring about how to use the exhibit and receiving helpful information from staff will have a more positive experience programming at Robot Park.

Table 73: Visitor Interactions Compared with Programming Rating (n=221)

Interactions with visitors	Rating Mean Rank		Sample Size		*U* value	*p* value
	Behavior Observed	Behavior not observed	Behavior Observed	Behavior not observed		
Overall	113.03	92.68	199	22	1786	.125
Operational	112.26	107.38	164	57	4467.5	.590
Positive Affect	122.46	84.08	155	66	3338.5	.000**
Goal Setting	122.24	98.58	116	105	4786	.003**
Conceptual	119.15	107.60	65	156	4540	.185
Negative Affect	103.49	113.43	54	167	4103.5	.282
Other	79.50	111.14	1	220	78.5	.593

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Table 74: Staff Interactions Compared with Programming Rating (n=221)

Interactions with staff	Rating Mean Rank		Sample Size		*U* value	*p* value
	Behavior Observed	Behavior not observed	Behavior Observed	Behavior not observed		
Overall	133.85	106.84	34	187	2402	.014*
Operational	133.21	107.38	31	190	2256.5	.024*
Positive Affect	143.55	109.46	10	211	729.5	.074
Goal Setting	136.83	109.12	15	206	1157.5	.079
Conceptual	135.81	109.45	13	208	1029.5	.118
Negative Affect	--	--	0	221	--	--
Other	122.17	110.85	3	218	293.5	.741

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Higher interest ratings were also related to higher programming ratings (See Table 75). Analyses found that respondents who were fairly interested in technology were significantly more likely to have higher

interest ratings for "programming the robot to move." Interest in "figuring out how computers work" did not affect programming ratings.

Table 75: Interest Ratings compared with Interaction Rating (n=221)

Interest Ratings	Rating Mean Rank		Sample Size		*U* value	*p* value
	Boring & Just okay	Pretty good & Exciting	Boring & Just okay	Pretty good & Exciting		
Learning about technology	90.79	113.65	35	184	2547.5	.034*
Figuring out how computers work	101.67	112.03	43	176	3426	.298

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Visitors to Robot Park will be less frustrated with the experience when Coach Mike is turned on. (Attitudes Indicator)

In order to determine Robot Park's difficulty level in regards to visitor experience, participants were asked "how easy was it to figure out what to do" at the exhibit. The median rating for difficulty was a 2 meaning that visitors perceived the exhibit was easy to figure out (Table 76). However, quite a few people found Robot Park a little difficult (38%) and 5% of visitors interviewed found the exhibit so difficult they couldn't figure it out, as seen in Table 77. Despite our hypothesis that visitors to Robot Park will be less frustrated with the experience when Coach Mike is engaged, there were no significant differences between treatment and control groups (Mann-Whitney U=6025, N=222, p=0.776).

Table 76: Difficulty of the Exhibit *(Scale:1="so easy I didn't have to think about it" and 4="so difficult I couldn't figure it out")*

How easy was it to figure out what to do here?	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	222	2.29	2	0.813	1	4
Treatment	115	2.30	2	0.818	1	4
Control	107	2.28	2	0.810	1	4

Table 77: Difficulty Ratings for the Exhibit

How easy was it to figure out what to do here?	Overall (n=222)	Conditions	
		Treatment (n=115)	Control (n=107)
So easy	18%	18%	18%
Easy	39%	37%	41%
Little difficult	38%	40%	36%
So difficult	5%	4%	5%

Additional independent variables were tested for relationships and there were several significant findings. Males found Robot Park easier to figure out than females (Mann-Whitney *U*=5103, N=222, *p*=0.027). Visitors who talked with other visitors about operational issues found the exhibit to be more difficult (Mann-Whitney *U*=3895.5, N=222, *p*=0.039). Also, visitors who made negative comments to other visitors found Robot Park significantly more difficult to figure out (Mann-Whitney *U*=3490, N=222, *p*=0.004). When interacting with staff, visitors who talked about goal setting were more likely to find Robot Park difficult (Mann-Whitney *U*=933, N=222, *p*=0.006). These findings suggest that visitors who have trouble working the exhibit are more likely to find Robot Park to be difficult.

Also, as average program length increased, visitors perceived difficulty level decreased (Ordinal Logistic Regression, *p*=0.038). Visitors who create short programs are more inclined to rate the exhibit as difficult.

Visitors with previous knowledge in technology and computers found Robot Park easier to figure out. Table 78 identifies the significant differences in knowledge ratings. Interest in technology or computers did not factor into visitors perception of exhibit difficulty (Mann-Whitney U test).

Table 78: Knowledge Ratings compared with Difficulty Rating (n=221)

Knowledge Ratings	Rating Mean Rank		Sample Size		*U* value	*p* value
	Nothing & A little	Quite a bit & A lot!	Nothing & A little	Quite a bit & A lot!		
Technology	124.64	100.91	94	127	4687	.004**
Computers	124.74	104.09	74	147	4422.5	.016*

Note. The Mann-Whitney test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Enjoyment of interacting with Mike AND Visitors will describe Coach Mike as either: a) a positive influence (supportive, helpful, motivating) or b) recognize his suggestions as helpful or useful (Attitude Indicator)

Visitors who interacted with Robot Park when Coach Mike was operating were asked additional questions during the interview about Coach Mike. The vast majority of those interviewed about Coach Mike indicated that they had noticed him (92%); these respondents were asked further questions about the nature of Coach Mike. When describing the role of Coach Mike, respondents tended to recall that he talked or communicated with them, usually giving general advice about the exhibit (30%), he acted as a teacher, coach, or guide to the exhibit (24%), or they used some variation of the word "help" in their description (i.e. helper, helpful, helped) (22%) (Table 79).

Additional analysis revealed that those who described Coach Mike as a teacher, coach, or guide were more likely have had no interactions with other visitors while at the exhibit (χ^2=4.275, N=106, p =.039); Coach Mike's role in scaffolding interactions was perhaps easier to recognize or more profound in the absence of other social interactions. Those who described Coach Mike as giving them specific advice were more likely to **complete the turn-around challenge** (χ^2= 7.894, 10.593, N=106, p = .005) and the **make-a-square challenge** (χ^2= 4.057, N=106, p = 0.044); they were also more likely to run a higher **number of programs** (Mann-Whitney U=245.5, N=106, p=.000) and **spend more time** at the exhibit (Mann-Whitney U=258.8, N=106, p=.001). These findings relative to describing Coach Mike as giving specific advice may be reflective of two alternate possibilities: 1) Coach Mike's ability to foster these results or 2) these respondents being "thorough users" of the exhibit, i.e. they had particularly in-depth experiences which included paying close attention to what Mike was saying.

Table 79: Respondents' Descriptions of the Role of Coach Mike (n=106)

Categories	n	Percent*
Communication-related and/or General Advice	32	30%
Teacher, coach or guide	25	24%
"Help"	23	22%
Specific Advice	13	12%
Positive Affect	10	9%
Negative Affect	8	8%
Computerized/Virtual	7	7%
I don't know/Not sure	6	6%
Other	9	9%

*Multiple responses allowed. Percentages total more than 100.

Respondents who remembered interacting with Coach Mike were asked specifically whether he helped them to "better understand the exhibit"; more than half (59%) said he was helpful, 38% indicated he was not helpful, and 3% were unsure.

- Of those who thought Coach Mike was helpful, 46% (n=30) gave general examples of how he was helpful; 49% (n=32) gave specific examples of how Coach Mike helped them understand the exhibit.
- Of those who felt Coach Mike was not helpful, 27% (n=8) did not listen to Coach Mike or did not feel they needed his help, 20% (n=6) said they experienced technical or usability issues, 20% (n=6) said Coach Mike was not helpful because he distracted them, and another 20% (n=6) felt Coach Mike make the exhibit more confusing or misleading.

Using the responses to the open-ended questions relative to the nature of Coach Mike, two variables were created to determine if evidence of the following indicators was found:

1. Visitors will enjoy interactive with the Virtual Humans. Very few participants (8%) who recalled interacting with Coach Mike received a score indicating positive enjoyment of their interaction with Coach Mike. Due to the small number of respondents, no further analysis using this variable was done.
2. Visitors will describe Coach Mike as either: a) a positive influence (supportive, helpful, motivating) or b) recognize his suggestions as helpful or useful. Two-thirds (66%) of all respondents described Coach Mike in helpful terms; this included individuals who did not feel that Coach Mike was helpful to them personally. Analysis of independent variables found that respondents who described Coach Mike as helpful were significantly more likely to have higher interest ratings for **"Figuring out how computers work"** (Mann-Whitney U=992.5, N=107, p=.041), higher ratings for **"Interacting with the exhibit"** (Mann-Whitney U=976.5, N=107, p=.029), and high ratings for **"Learning more about computers by interacting with the exhibit"** (Mann-Whitney U=972.0, N=107, p=.032).

Participants' Attitudes towards Coach Mike on the Follow-up Online Questionnaire

Participants who interacted with Robot Park when Coach Mike was operating were asked additional questions in the follow-up questionnaire about Coach Mike. The majority of respondents indicated that they remembered Coach Mike (75%). These respondents were then asked how Coach Mike helped them "to better understand how to program the robot at Robot Park." The majority (74%) found Coach Mike helpful. Two respondents (9%) did not recognize Coach Mike's suggestions as helpful or useful. When describing the role of Coach Mike, respondents tended to recall that he gave general advice about the exhibit (63%), gave specific advice or directions (31%), or they used some other description for Coach Mike (13%) (Table 80). Some examples of these responses include:

Coach Mike told us what to do and the object of the exhibit.
Coach Mike helped me better understand how to begin the chain of movements.
He tried to explain to line up the action blocks and then hit the button that would photograph the list of actions we wanted.
It made it more entertaining, not just a science experiment

Table 80: Helpful Nature of Coach Mike (n=16)

Responses	n	Percent*
General advice/directions	10	63%
Specific advice/directions	5	31%
Other	2	13%

*Multiple responses allowed. Percents may add up to more than 100.

Coach Mike and Robot Park: Visitor Awareness

Table 81 shows the Awareness impacts and indicators for Coach Mike and the Robot Park exhibits. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 81: Coach Mike Awareness High-Level Results

Impact (Shaded) and its Related Indicators	Level of Evidence Based on the Summative Evaluation
Children (ages 7 -14) and adults will increase their awareness about computer science and technology.	*Achieved: 1 of 1 indicators showed evidence of increased awareness.*
Visitors will recognize at least one of the following as **characteristics of Coach Mike**: a) his responses are tied to what the visitor is doing, b) he is trying to help them problem-solve.	• 73% who interacted with Coach Mike recognized at least one of the characteristics named in the indicator. • The most common response was related to the helpfulness or teacher/coach/guide-like qualities of Coach Mike, the next common response was problem-solving-related answers, and finally that his responses were tied to what the participant was doing.

Visitors will recognize at least one of the following as characteristics of Coach Mike: a) his responses are tied to what the visitor is doing, b) he is trying to help them problem-solve. (Awareness Indicator)

Using the responses to the open-ended questions relative to the nature of Coach Mike, three variables were created to determine if evidence of the following indicators were found: *Visitors will recognize at least one of the following as characteristics of Coach Mike: a) his responses are tied to what the visitor is doing, b) he is trying to help them problem-solve.* The results of the analysis of these variables are reported below:

- Variable for Coach Mike's "responses are tied to what the visitor is doing." Very few participants' open-ended responses (6%, n=7) demonstrated that they understood Coach Mike's responses are tied to their actual actions and not randomly generated responses. Although the number was small, these visitors seemed to recognize that Coach Mike was directly responding to their actions when he would make suggestions directly related to something they did incorrectly, such as physically getting in the way of the camera. This finding indicates that users only realize Coach Mike's responses are not random when they clearly make an obvious error, and Mike highlights that error.
- Variable for Coach Mike "is trying to help [the visitor] problem-solve." The majority of participants' open-ended responses (72%, n=82) included a statement that could be indicative of Coach Mike as helping with problem-solving. The majority of these comments were related to Mike as a teacher, coach, or guide or his helpful nature (87%, n=71). Fewer respondents specifically used terms equated specifically with problem-solving (13%, n=11); however, given the open-ended nature of question, it would not be expected that visitors would use the same language as the indicator.
- Variable for Awareness of Coach Mike's Characteristics. The two proceeding variables were combined to create one overarching awareness variable. By this measure, the majority of respondents (73%, n=83) who interacted with Coach Mike recognized at least one of the characteristics.
    - When additional analysis was performed, independent variables such as age (adult/child), time spent, number of programs created, length of programs, knowledge and interest in computers and technology yielded no significant differences. Interactions with Coach Mike in terms of completing challenges and observable behaviors (i.e. looking at Mike, talking to Mike, mentioning Mike, etc.) were also not significant.
    - One of the three exhibit-related rating statements was significant; Visitors with higher ratings of **"Learning more about computers by interacting with the exhibit would be…"** were more likely to demonstrate awareness of the characteristics of Coach Mike defined in the indicator (Mann-Whitney *U*=973.0, N=114, *p*=.035).

Participants' Awareness of the Goals of the Robot Park Exhibit

Although it was not an indicator of awareness as represented by the Impacts and Indicators, participants' recognition of the main idea of the exhibit was an additional measure of awareness. When asked what they thought the main idea was for Robot Park, visitors most frequently mentioned computer programming (28%), followed by increased knowledge or awareness about robots (26%), computers, science or technology more generally (19%), completing a specific task (13%), positive attitudes towards computers/robots (5%), and to stimulate curiosity and interest (5%). (See Table 82 for a summary). Some examples of these responses include:

To teach children the idea of putting commands together and how robots read commands.
To teach people how robots work.
To learn about computer science.
To try to get the robot to the target area.
Get people thinking.

Table 82: Responses for "What would you say was the main idea of the exhibit?"

Categories	Overall Exhibit (n=222)		Treatment (n=115)		Control (n=124)	
	n	Percent*	n	Percent*	n	Percent*
Computer programming	61	28%	31	27%	30	28%
Robots	57	26%	30	26%	27	25%
Computers, science or technology in general	41	19%	23	20%	18	17%
Completing a specific task	29	13%	19	17%	10	9%
Positive attitudes about computers/robots	12	5%	5	4%	7	7%
Stimulate curiosity & interest in general	5	2%	4	4%	1	1%
I don't know	23	10%	11	10%	12	11%
Other	15	7%	6	5%	9	8%

* Multiple responses allowed. Percents may add up to more than 100.

There were no significant differences between how the treatment group and control group responded to this question (Chi-Square tests). However, age group did make a difference in responses, as shown in Table 83. Adults were significantly more likely to cite the main idea as related to computer programming than children. On the other hand, children were more likely to respond that Robot Park helped visitors increase knowledge and awareness about robots. Children also were more inclined to generally mention computers, science or technology. Adults responded with positive attitudes towards computers or robots significantly more than children. Children also demonstrated a significantly higher percentage of "I don't know" responses. Overall, adult responses were more focused on ideas related to programming and positive experiences with robots—more abstract concepts, while children saw the main idea of Robot Park as learning about robots or task-based—more concrete concepts.

Table 83: Main Idea by Age group

Responses	Adults (n=74)		Children (n=148)		χ^2 value	*p* value
	n	Percent in group	n	Percent in group		
Computer programming	34	46%	27	18%	18.999	.000**
Robots	12	16%	45	30%	5.205	.023*
Computers, science or technology	8	11%	33	22%	4.323	.038*
Completing a specific task	5	7%	24	16%	3.887	.058
Positive attitudes	8	11%	4	3%	6.343	.012*
Stimulate curiosity & interest	3	4%	2	1%	n/a	n/a
I don't know	3	4%	20	14%	4.753	.029*
Other	6	8%	9	6%	0.307	.580

Note. The Pearson Chi-Square test was used to test for statistical significance. * Significant at the p<.05 level. **Significant at the p<.01 level.

Responses were also analyzed by stay time, number of programs, and length of programs. Several significant findings emerged. Visitors who described the main idea as increasing knowledge and awareness about robots created more programs at Robot Park (Mann-Whitney U=3694.5, N=222, p=.015). Visitors who discussed the main idea of Robot Park as a way to generally stimulate curiosity and interest created significantly less programs than visitors who did not mention this as part of the main idea (Mann-Whitney U=244.5, N=222, p=.035).

Interest ratings also factored into visitors' responses. Visitors with little to no interest in technology were more inclined to respond with "I don't know" for the main idea of Robot Park than visitors with a pretty high interest in technology (χ^2= 18.578, N=220, *p*=0.000). Visitors who found learning about technology to be "boring" or "just okay" also had a higher frequency of "I don't know" responses.

Coach Mike and Robot Park: Visitor Knowledge

Table 84 shows the Knowledge impacts and indicators for Coach Mike and the Robot Park exhibit. For each impact and indicator a summary of the evidence is provided. Detailed explanations of the findings for each indicator are presented in the sections following the table.

Table 84: Coach Mike Knowledge High-Level Results

Impact (Shaded) and its Related Indicators	Level of Evidence Based on the Summative Evaluation
Children (ages 7 -14) and adults will increase their knowledge about computer science and technology.	*Not achieved: 1 of 1 indicator showed no differences between the control and treatment.*
Visitors to Robot Park will demonstrate an **increase in their knowledge of programming concepts** (i.e. the nature of programming) when Coach Mike is turned on, as indicated at least one of the following: a) programming as a series of steps, b) programs are rarely correct the first time, c) revisions are often necessary, and d) failure is not the end of the process.	• Onsite open-ended analysis indicated that 27% of participants identified at least one of the concepts. There was no difference between conditions. • Using onsite ratings given by adults, self-reported knowledge of programming concepts increased after interacting with the exhibit. There was no difference between conditions. • Follow-up data indicate that 65% of respondents were able to correctly categorize statements pertaining to programming concepts. There was no difference between conditions.
Visitors can describe that **computers are programmed by breaking large programs/tasks into smaller steps**.	*(Subsumed by the indicator above)*

Visitors to Robot Park will demonstrate an increase in their knowledge of programming concepts (i.e. the nature of programming) when Coach Mike is turned on, as indicated at least one of the following: a) programming as a series of steps, b) programs are rarely correct the first time, revisions are often necessary, failure is not the end of the process and c) users are needed/robots can't program themselves. (Knowledge Indicator)

A series of quantitative variables were created to address this indictor. Participants' responses throughout the entire interview were reviewed to create variables that addressed each programming concept. These variables were then used to create a three-point rubric to address the indicator as a whole. Adult participants' responses to two retrospective-pre/post/delayed post rating questions were also used as evidence of this indicator.

Participants' Knowledge of Programming Concepts

Visitors were asked three open-ended questions where they could relate their knowledge of programming concepts: 1) "Can you tell me what you did at the exhibit?"; 1a) "Did anything you just did help you understand how computers work?"; and 2) "What would you say was the main idea of the exhibit?" Combining responses from the open-ended questions regarding programming concepts, three variables were created to determine knowledge gain relative to programming.

1. Computers are programmed through small steps: During the course of the interview, participants described at least one of the following concepts related to the programming: a) programming as a series of steps, b) programs have an "order" or "sequence," and c) computers are programmed by breaking large programs/tasks into smaller steps.
2. Trial and Error: Participants mention during their interview the concept that a) programs are rarely correct the first time, b) revisions are often necessary, and c) failure is not the end of the process.[13]
3. User as programmer: Participants' responses during the interview include the concept that a) a user (i.e. a person) programs and controls a robot or b) a robot needs a program or cannot operate without a program.

About one third of the visitors interviewed (31%) mentioned at least one of the defined programming concepts (Table 85). One out of every five visitors (19%) demonstrated increased knowledge that computers are programmed through small steps, followed by 9% of visitors recognizing that a user programs and controls a robot, and 3% of visitors expressing the idea of programming through trial and error. There were no significant differences between condition groups. There were, however, significant differences between adults and children. Adults more frequently cited programming as a series of small steps (Mann-Whitney *U*=4152, N=223, *p*=.000).

Table 85: Programming Concepts

Concepts	Overall Exhibit (n=223)		Treatment (n=115)		Control (n=124)	
	n	Percent*	n	Percent*	n	Percent*
Small steps	43	19%	22	19%	21	19%
Trial and error	7	3%	2	2%	5	5%
User as programmer	21	9%	11	10%	10	9%

*Multiple responses allowed. Percents may add up to more than 100.

When comparing programming concept comprehension with time, number of programs and average program length, one significant relationship emerged. Visitors recognizing that a human is necessary to program and control a robot created significantly more programs (Mann-Whitney *U*=1504, N=223, *p*=.028). The more programs a visitor creates, the more likely they are to verbalize that at the heart of every robot is a computer program, written by a person.

Overarching Rubric of Participants' Knowledge of Programming

Responses to these questions were combined into a three-point coding rubric:

[13] Although these were three separate concepts in the indicator, it proved too difficult during the coding of the interviews to distinguish between these concepts in visitor responses.

- No understanding (0 points)**:** Participants' responses do not include any of the concepts describe above: a) computers are programmed through small steps, b) trial and error, or c) user as programmer.
- Basic understanding (1 point): Visitor names one of the identified programming concepts.
- Advanced understanding (2 points): Visitor names two or more programming concepts.

The mean knowledge level using this rubric was 0.4 or "no understanding". Table 86 presents the distribution of results across the scale, and show that the majority of visitors demonstrated no understanding of programming concepts. By combining the basic and advanced categories for the purposes of addressing the indicator, a total of 27% of participants recognized at least one programming concept. Researchers compared understandings between the treatment and control groups. There were no significant differences between the two groups.

Table 86: Understanding of Programming Concepts

Understanding	Overall Exhibit (n=223)		Treatment (n=115)		Control (n=124)	
	n	Percent	n	Percent	n	Percent
No understanding	163	73%	85	74%	78	72%
Basic	50	22%	25	22%	25	23%
Advanced	10	5%	5	4%	5	5%

However, age did contribute to visitors understanding of programming concepts. Adult visitors (16 years and older) demonstrated more advanced understandings of programming concepts than children did (Mann-Whitney U=4348.5, N=223, p=.001). There is also a relationship between stay time and understanding of programming concepts (Ordinal Logistic Regression, p=0.031). As stay time increases, the chance of visitors comprehending more advanced computer programming concepts also increases.

Adults' Rating of their Programming Knowledge

A separate quantitative approach was used with adults to determine if interacting with the exhibit impacted self-reported agreement with 1) explaining "what it means to write a computer program" and 2) knowing "what it means to 'debug' a program." When looking at adults' ratings overall, regardless of condition, adults had a significantly higher ratings of both measure of programming knowledge after their interaction with the exhibit (See Table 87).

Table 87: Adult Respondents' Ratings (Retrospective-Pre/Post) of their Programming Knowledge *(Scale: 1="strongly disagree," 7="strongly agree"*; n=73)

Statement	Retrospective-Pre Rating (Mean)	Post Rating (Mean)	*Z value*	*p* value
I can explain what it means to write a computer program.	3.3	4.1	-4.315	.000**
I know what it means to "debug" a program	4.1	4.5	-3.493	.000**

Note. The Wilcoxon Signed Ranks test was used to test for statistical significance.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

When the results were analyzed based on condition, both treatment and control participants showed statistically significant changes in their ratings retrospective-pre to post. This indicates that for these measures, Coach Mike neither supported nor hindered the meaning-making regarding programming that was occurring at Robot Park. The significance levels for each statement by condition are reported below:

- *Treatment: I can explain what it means to write a computer program.* Significant change in mean rating retrospective-pre to post (*Z*=-2.818, N=28, *p*=.005).
- *Control: I can explain what it means to write a computer program.* Significant change in mean rating retrospective-pre to post (*Z*=-3.334, N=45, *p*=.001)
- *Treatment: I know what it means to "debug" a program.* Significant change in mean rating retrospective-pre to post (*Z*=-2.598, N=28, *p*=.009).
- *Control: I know what it means to "debug" a program.* Significant change in mean rating retrospective-pre to post (*Z*=-2.401, N=45, *p*=.016)

When comparing the onsite ratings with the online questionnaire ratings (i.e. post to delayed post), there were no significant differences in how either of the statements were rated. In other words, respondents' perceptions of their knowledge held constant in the six weeks after the museum visit. This could be interpreted as the gains that were made directly after interacting with the exhibit were maintained overtime. Another possibility is that these results are a "practice effect" from answering the same question multiple times. Also possible is that not enough time had passed for a decay of the effects to have occurred. A questionnaire conducted six months after having used the exhibit may yield different results. There were no differences between the ratings of the treatment and control groups in how the statements were rated in the delayed post.

Participants' Knowledge of Programming based on the Follow-up Online Questionnaire

The goal of Robot Park is to help visitors learn about programming. One question in the follow-up questionnaire was designed to test respondents' ability to describe programming. Respondents were provided with the following list of statements:

- A series of small steps
- Combining small steps into something bigger
- There is usually only one way to solve a problem
- You might not be right the first time
- You might have to "debug" your program

- If you mess up, you can't fix it

Using their knowledge about programming, respondents were asked to place each statement into one of two boxes: 1) Yes, this describes programming; 2) No, does not describe programming. As shown in Table 88, the majority of respondents (65%) placed all six statements in the right box, correctly identifying the true and false statements regarding programming. "You might have to 'debug' your program" and "If you mess up, you can't fix it" had the highest percentage of incorrect answers, with 7 respondents (11%) improperly identifying these statements (Table 89). Overall, respondents seem to recognize computer programming concepts.

Table 88: Number of Correctly Identified Statements about Programming (n=65)

Correctly Identified	n	Percent
No statements	0	0%
1 statement	0	0%
2 statements	1	2%
3 statements	2	3%
4 statements	8	12%
5 statements	12	19%
All statements	42	65%

Table 89: Frequency of Correctly Identified Statements (N=65)

Correctly Identified	Number of people	n	Percent
A series of small steps	59	64	92%
Combining small steps into something bigger	59	63	94%
There is usually only one way to solve a problem	60	63	95%
You might not be right the first time	61	63	95%
You might have to "debug" your program	55	62	89%
If you mess up, you can't fix it	58	65	89%

There was no significant difference between treatment and control groups (Mann-Whitney U=400.5, N=65, p=0.054). Visitors to Robot Park were able to grasp concepts of programming with or without Coach Mike engaged. Age group (adult vs. child), gender, stay time, number of programs, or length of programs also did not affect respondents' ability to correctly identify programming statements.

The follow-up questionnaire responses reflect a greater understanding of computer programming concepts than the on-site exhibition interview. This is expected as scores on true-false items tend to be high due to the ease of guessing correct answers when the answer is not known. Also, the on-site interview asked a series of open-ended questions in order to determine visitors' understanding of programming concepts. An open-ended question is designed to encourage a full, meaningful answer using the subject's own knowledge and/or feelings. Open-ended questions also tend to be more

objective and less leading. However, this may be taken as evidence that participants may have had higher levels of programming knowledge than the onsite analysis indicated.

Increase is STEM Knowledge as a Result of Interacting with Robot Park

Participants were asked both onsite and online questions relating to what they thought they had learned about computers through interacting with the exhibit. The onsite replies were included in the overarching interview analysis of knowledge of programming concepts. Both sets of responses are included below.

Participants' Self-reported STEM Learning during the Onsite Interview

During the interview, visitors were asked whether anything they just did helped them "understand how computers work." About half (48%) said the exhibit was helpful, and 53% felt the exhibit did not help them understand more about computers. Visitors attending Robot Park with Coach Mike engaged were significantly more likely to find that the exhibit helped them understand how computers work (χ^2= 3.875, N=223, *p*=0.049). Of those that found Robot Park did help them understand how computers work, the majority said that the exhibit helped them to understand how a program is read by a computer and transmitted to the robot (36%) (Table 90). Representative responses include:

Yeah-like the coding on the blocks what's read by a computer and transmitted.
It was interesting observing the pieces and what was on them and how the computer read them to tell the robot what to do.
Yes: that there was circles on pieces of puzzle that computer read them and he had a different name for them that the robot read to work.

This was followed by visitors recognizing that the wooden blocks represented actions for the robot to perform (21%). One respondent said "to me it's actually pretty cool that when you put the blocks together you can make a robot do things." Fifteen percent of visitors mentioned the theme that a person does the programming and is necessary to control the robot.

Table 90: Respondents' Descriptions of how the Exhibit helped Visitors Understand more about Computers (n=106)

Responses	n	Percent*
Program read by computer and transmitted to robot	38	36%
Blocks represent actions	22	21%
User programs and controls robot	26	15%
Computers follow a series of steps	11	10%
Write a computer program	10	9%
Other	26	25%

*Multiple responses allowed. Percentages total more than 100%

When each category of response was compared by condition (treatment or control), significant differences were discovered. Visitors to Robot Park without Coach Mike engaged were significantly

more likely to say that blocks represent actions for the robot to perform and that computers follow a series of steps (See Table 91). When comparing between age groups (adults and children), children were significantly more likely to say that their experience at Robot Park helped them understand how computers work, than were adults (χ^2= 7.502, N=223, p=0.006). Adults, on the other hand, were significantly more likely to realize that computers follow a series of steps (χ^2= 5.974, N=106, p=0.015).

Table 91: Understand more about Computers by Condition

Categories	Treatment (n=115)		Control (n=108)		χ^2 value	p value
	n	Percent in group	n	Percent in group		
Program read by computer and transmitted to robot	21	34%	17	39%	0.254	.614
Blocks represent actions	8	13%	14	32%	5.598	.018*
User programs and controls robot	10	16%	6	14%	0.125	.724
Computers follow a series of steps	3	5%	8	18%	4.927	.026*
Write a computer program	7	11%	3	7%	0.602	.438
Other	17	27%	9	21%	0.674	.412

Note. The Pearson Chi-Square test was used to test for statistical significance. Fisher's exact test was also applied and yielded the exact same results.
* Significant at the p<.05 level. ** Significant at the p<.01 level.

Significant differences were also discovered when each category of response was compared with social interactions. Visitors who interacted with other visitors more frequently found the exhibit helpful in understanding how computers work (χ^2= 7.502, N=223, p=0.006), especially visitors who spoke with other visitors about completing or setting goals (χ^2= 4.452, N=223, p=0.035). Visitors made a positive remark or comment to other visitors were significantly more likely to find the exhibit helpful in understanding how computers work (χ^2= 10.067, N=223, p=0.002). When comparing responses with staff interactions, visitors who discussed goal setting with staff members more frequently found the exhibit helpful for understanding computers (χ^2= 6.797, N=223, p=0.009).

Researchers also examined the relationship between stay time, number of programs and program length with responses. The following significant findings emerged. Visitors who responded that their experienced help them recognize that they were writing a computer program were more likely to spend longer times in the exhibit than those that did not mention writing a computer program (Mann-Whitney U=285.5, N=106, p=.036). Regarding number of programs, visitors who felt that Robot Park helped them understand how computers work were more likely to create more programs (Mann-Whitney U=4743.5, N=223, p=.002).

Knowledge of computers also factored into visitor responses. , Visitors with little knowledge of computers were significantly more likely to mention that computers follow a series of steps than those

visitors pretty knowledgeable about computers (χ^2= 13.216, N=106, p=0.000). This is an interesting finding as visitors with pre-existing knowledge of computers might have already understood that computers follow a series of steps, while visitors with little knowledge of computers learned this through the Robot Park exhibit.

Participants' Self-reported STEM Learning on the Follow-up Online Questionnaire

In order to address STEM learning, respondents to the online questionnaire were asked what they learned "about technology or computers from interacting at Robot Park." About a quarter of responses (26%) generally mentioned the exhibit, robots or technology (Table 92). Some respondents (22%) recognized that computers follow a series of steps, require code, or a sequence of commands.

Computers use sequences of codes to operate.
I learned that robots operate by following a certain code of movements.
I already knew that computers take a series of commands and only do exactly what they say.

Fourteen percent of respondents felt that the exhibit helped them to understand how a program is read by a computer and transmitted to the robot. This was followed by 12% articulating that a person is necessary to program and control a robot. One respondent said "when you build commands robots can obey humans." A few respondents (6%) mentioned that programming can be difficult and that there is a chance for "human error."

Table 92: STEM learning at Robot Park (n=64)

Responses	n	Percent*
General comments about exhibit, robots or technology	13	26%
Computers follow a series of steps	11	22%
Blocks read by computer and transmitted to robot	7	14%
User programs and controls robot	6	12%
Programming can be difficult and there might be errors	3	6%
Don't know/nothing new	8	16%
Other	6	12%

* Multiple responses allowed. Percents may add up to more than 100.

There were no significant differences in responses for treatment and control groups. When each category of response was compared by age group (adult or child), one significant difference was discovered. Adults were significantly more likely to say that their experience at Robot Park helped them understand that computers follow a series of steps (χ^2=4.675, N=50, p=0.031).

Researchers also compared overall museum ratings to individual responses. Respondents who mentioned learning that computers following a series of steps rated the educational experience of the Museum of Science higher than those that did not discuss that computers follow a series of steps (Mann-Whitney U=133.5, N=50, p=0.005). Learning at Robot Park influences the overall educational experience at the Museum of Science.

Conclusions

The summative evaluation of the *Responsive Virtual Humans Museum Guides* project demonstrated mixed results in incorporating virtual humans into the museum environment. Generally, the Twins and Science Behind (taken together) were successful at achieving the impacts the team identified for the exhibits. The study of Coach Mike, on the other hand, demonstrated that the exhibit was only somewhat successful in achieving the identified impacts. Returning to the overarching evaluation questions provides a summary of the findings at a high level.

1. **What is the nature of visitors' interactions with the three *Virtual Human* exhibits?**
    a. *Who uses the exhibits?*
    Using the target visitor who was observed to answer this question, children were more likely than adults to be the first individual in their group to approach and engage with the exhibits. This finding is in keeping with what has been observed by staff at the museum, with the family's experience strongly shaped by where the children choose to stop.
    b. *How do visitors approach and move between the Twins and the Science Behind?*
    This study shows that visitors did not frequently move from the Twins to Science Behind, as was originally intended by the project team. The initial idea was that visitors would move naturally from the Twins to the Science Behind to learn more about how the Twins worked; however, this study shows that very few visitors demonstrate that behavior. No summative data was collected on visitor movements from Science Behind to the Twins; in the piloting stage, this type of movement was a rare occurrence and was therefore eliminated as a focus of the summative evaluation.
    c. *What are the differences in visitors' interactions with the Twins (types of questions, number of questions, stay time, social interactions) when it is staffed by an MOS interpreter (Blended interaction) verses when it is unstaffed and visitors are interacting with the Twins directly (Direct interaction)?*
    There were clear differences between the Blended and Direct interaction approaches tested at the Twins. This is not to say that one type of interaction is better than the other; rather, each approach has unique strengths. This finding supports a flexible facilitation approach, where staff at Cahners Computer Place can choose to incorporate either Blended or Direct interaction depending on their staffing levels and which outcomes they are looking to support. Specific findings indicate that in the Blended condition, visitors were significantly more likely to: 1) discuss personal and technology/computer-related topics; 2) stay longer at the exhibit (nearly 3 minutes longer); and 3) interact with MOS staff. In the Direct condition, visitors were significantly more likely to interact with other visitors. The facilitation approach (whether Blended or Direct) did not influence the number of verbal utterances visitors addressed to the Twins.
    d. *What are the differences in visitors' interactions with Robot Park (stay time, number of programs created, types of programs created, completion of specific tasks, social interactions) when the virtual human guide (Coach Mike) is present and when he is not present?*
    Overall, the presence of Coach Mike at Robot Park did not fully achieve the hypothesized impacts. No significant statistical differences were discovered between the number of programs created, program length, or social interactions when the two conditions were compared. However, some visitor behaviors were influenced positively

by the presence of Coach Mike. For example, visitors who engaged with the exhibit with Coach Mike did have significantly longer stay times than those who visited with exhibit without Coach Mike. Coach Mike also supported visitors in completing one of the exhibit challenges (programming the robot to illuminate a sign), as visitors who interacted with Robot Park when Coach Mike is engaged were significantly more likely to complete that goal than are visitors who do not have Coach Mike support. Coach Mike's presence also decreased undesirable behaviors such as 1) using the block tester to move the robot and 2) the writing of long programs instead of editing more, shorter programs. Taking these findings in conjunction with results that show that the current mechanism used to trigger Coach Mike leads to increased usability issues and the highly successful nature of Robot Park without Coach Mike, it appears that the value of Coach Mike is limited to these behavioral improvements, at least in its current iteration.

2. **In what ways do interactions with the *Virtual Human* exhibits impact visitors' knowledge and awareness of, engagement and interest in, and attitudes and perceptions towards computer science and technology?**

   The Twins and Science Behind (taken together) positively impacted visitors across all four impact categories: Engagement and Interest, Attitudes, Awareness, and Knowledge. Overall, visitors found the exhibits engaging, whether the Twins are staffed or unstaffed. Visitors recognized that The Twins are examples of virtual humans and were easily able to identify features of virtual humans. Visitors were curious about the Twins and their capabilities and hold generally positive views about virtual humans in society. As a set of stand-alone exhibits highlighting an advancement in technology and computing, the Twins and Science Behind are successful.

   Coach Mike, on the other hand, was not as successful in achieving visitor impacts in the same four areas. This study found evidence for five out of the eleven indicators identified as impact measures; Coach Mike did impact visitors Awareness, Engagement and Interest, and Attitudes for some of the indicators defined for each of these impacts. There was no evidence of a Knowledge impact using the identified indicators. Despite a lack of statistically significant differences between measures such as exhibit ratings, number of programs created, or program length, some visitors who experienced Coach Mike perceived that the intelligent tutor was helpful as a teacher, coach, and guide. There was also evidence that the presence of Coach Mike might have helped visitors to work with the TERN programming language, significantly reducing behaviors such as using the block tester for the majority of movements, pushing run without the start block, and pushing run without creating a program.

   There are several ways these results can be interpreted. First, it is important to consider that the Robot Park exhibit existed at MOS before the current project ever took shape. Coach Mike was envisioned as an additional feature at Robot Park that would enhance visitor outcomes at the exhibit. Even before the arrival of Coach Mike, other studies found that Robot Park was a successful, well-designed, intuitive exhibit. Therefore, improving such a successful exhibit is a difficult task. It may be that given the overall high performance of Robot Park, there is little room for overall improvement at the exhibit. Second, usability issues at Robot Park were prevalent when Coach Mike was engaged throughout the duration of this study; had these issues been resolved, visitor impacts relative to Coach Mike might have been stronger. Third and perhaps most important for future work, Coach Mike represents the first known use of an intelligent tutor in an informal education setting. The project team was just beginning their investigation into how this novel addition to a museum might perform; as a result, some of the indicators defined by the team were demonstrated by the evaluation to be unlikely given how

visitors interacted with the exhibit and Coach Mike. As the team continues to investigate the integration of virtual humans and intelligent tutors into informal education settings, they will need to continually refine their expectations for what constitutes "success" in these settings. For example, behaviors or measures that are evidence of success in a formal education setting may not transfer to an informal setting. Further studies, such as the one currently being conducted by MOS evaluation staff, may reveal additional ways of measuring success for these interactions or may add to the team's understanding of how museum visitors respond to virtual humans and intelligent tutors.

Submitted: April 25, 2012

Appendices

Appendix 1 **Protocol**

Protocol for the Summative Evaluation

June 16, 2011

Prepared by:
Susan Foutz, MA
Senior Research Associate

Introduction

The evaluation team at the Institute for Learning Innovation (ILI) seeks IRB approval for the summative evaluation of the project *Responsive Virtual Human Museum Guides.* The University of Southern California's Institute for Creative Technologies (ICT) and the Museum of Science, Boston (MOS) are the lead institutions for the project with funding provided by an Informal Science Education grant from the National Science Foundation (#08133541). The project seeks to collaboratively create a life-sized 3D virtual, computer generated character to serve as a museum "educator" at MOS.

In this, the third year of the project, the summative evaluation seeks to determine to what degree the project has met its public impact goals for the project. The summative evaluation is lead by ILI staff member Susan Foutz, Senior Research Associate. (Contact information: Institute for Learning Innovation, 3168 Braverton Street, Suite 280, Edgewater, MD, 21037. Phone: 410-956-5144. Email: foutz@ilinet.org.) Additional members of the team include ILI Research Associate Jeanine Ancelet and Emily Shapero, a contract evaluator with ILI.

The *Year 3 Summative Evaluation Work Plan* is a separate document in this submission that includes the rationale for the project, evaluation questions, and the study design. This document, the *Protocol for the Summative Evaluation*, details the procedures for implementing the evaluation plan, specifically issues of recruitment, sampling, obtaining consent, relevant instruments, data collection, data analysis, reporting, and data storage where human subjects are involved.

The summative evaluation focuses on two experiences in *Cahners Computer Place* (CCP) at MOS: 1) The Twins, Ada and Grace, and the accompanying exhibit *The Science Behind*, and 2) Coach Mike at Robot Park. See the *Year 3 Summative Evaluation Work Plan* for details on the experience. Each experience will be evaluated using a combination of visitor observation in CCP, visitor interviews, and follow-up web questionnaires. The remainder of this document is organized by the two experiences that serve as the focus of the summative evaluation.

The Twins, Ada and Grace, and The Science Behind

The summative evaluation will include measures of visitor interaction and impacts at the Twins and The Science Behind. Originally thought of as two separate but related exhibits, the exhibits are in fact very closely related; visitors move easily back and forth between the exhibits in part due to suggestions made by Twins. As such, it was decided to treat these two exhibits as one experience for the summative evaluation. The evaluation of the Twins and The Science Behind will include visitor observation, visitor interviews, and a follow-up online questionnaire. The instruments to be used and the consent, sampling, recruitment, and incentive procedures for each of these methods is detailed below.

A. Visitor Observations

- *Instrument:* Observations of visitors to the Twins and The Science Behind will be recorded using the *Twins Observation* Instrument. This instrument will be used to record questions and utterances visitors make to the Twins, the Twins' responses, interaction between the target visitor and other visitors and MOS staff, easily observed behaviors and interactions at The Science Behind, and any usability issues. See Appendix A for the guide to how visitor behaviors will be recorded on the *Twins Observation* Instrument.
- *Passive Consent:* While ILI staff members are present in CCP and conducting observations, a sign notifying visitors of the observation will be posted near the Twins and The Science Behind exhibits. The sign will read: "To help us better serve our visitors, this area is being observed ending at [time]. With any questions or concerns, please contact Christine Reich at 617-589-0302."
- *Sampling:* The sampling frame for the observations includes all visitors who appear to be at least 7 years old and who interact with the Twins exhibit. For the purposes of this observation, "interaction" with the exhibit is defined as pushing (or attempting to push) the button on the microphone and directing at least one question or utterance to the Twins. In order to select the "target" visitor who will be the focus of an observation, ILI staff will use the following method:
    - When the data collector is ready to begin an observation (i.e. immediately after a break or after completing an interview), they will station themselves within 5 to 10 feet of the Twins microphone with a direct line of sight to the microphone.
    - The next visitor who appears to be at least 7 years old and who 1) plants both feet in front of the podium on which the microphone is mounted or 2) sits on the stool will be the target for the observation and timing will start. If the target does not interact with the exhibit (as defined above), the data collector will observe the next available visitor who meets the above criteria, restarting the timing.
    - The observation will continue as long as the target visitor is 1) interacting with the Twins, 2) watching another visitor interact with the Twins, or 3) engaging with The Science Behind. Once the visitor either 1) moves to exit CCP or 2) begins to engage with another exhibit, the data collector will approach the target and request an interview.
- *Recruitment:* No recruitment method will be used with visitors before beginning the observations.
- *Incentive:* There are no incentives given to visitors in exchange for their participation in the observations.

B. Visitor Interviews

- *Instrument:* Interviews with visitors to the Twins and The Science Behind will be guided by the *Twins Interview* Instrument. This instrument will be used to ask visitors open-ended and rating scale question about 1) their interactions with the Twins and The Science Behind, 2)changes in their knowledge, attitudes, awareness, and engagement as a result of their interactions, 3) basic demographic and psychographic information, and 4) an email address for the purposes of contacting visitors for a follow-up web questionnaire. The interview will take approximately 8 to 10 minutes to complete.
- *Sampling:* The sampling frame for the interviews includes all visitors who were targets of an observation, with the goal to create a matched observation of and interview with the same target visitor. After the observation of a target visitor ends (as defined in the Visitor Observation section above), the ILI staff member will approach the target visitor for an interview. While other members of the target visitor's group may participate in the interview, the data collector will ensure the target answers all questions and supplies their demographic and psychographic information.
- *Consent and Recruitment:* After the observation of a target visitor ends, the ILI staff member will approach the target visitor for an interview. The script included at the beginning of the *Twins Interview* instrument will serve as both the consent and recruitment language for visitors. Visitors who agree to speak with the interviewer have consented to participate in the interview. Those visitors who appear to be under 16 years of age will be asked for their assent to participate in the interview *after* the adult who is supervising them gives their consent to the interview. In the event that a target visitor declines to participate in an interview (or to have their child under 16 participate in an interview), the visitor will be thanked by the ILI staff member and allowed to continue their visit; the observational data from visitors who decline to participate in an interview will be maintained by ILI project staff and analyzed 1) as a point of comparison to those visitors who did consent to the interview and 2) to increase the overall sample size of the observations.
- *Incentive:* In thanks for completing the interview, visitors will be offered a small thank-you gift with a value of less than $1, such as a pencil.

C. Follow-up Questionnaire

- *Instrument:* A follow-up online questionnaire with target visitors will be guided by an online version of the *Twins Follow-up Questionnaire* Instrument. This instrument will be used to ask visitors open-ended and rating scale question about their knowledge, attitudes, awareness, and engagement outcomes as a result of their interactions with the Twins and The Science Behind at MOS and after they left the museum. The questionnaire will take approximately 5 minutes to complete. The online questionnaire will be hosted by Qualtrics, a program with which ILI researchers have extensive experience. Qualtrics allows the ILI project staff to create and maintain the online questionnaire in a password-protected environment accessible only to the ILI project staff. This maintains the safety of the data while it is within the Qualtrics system. Once all questionnaire data has been collected, the data will be downloaded to ILI's secure server. Questionnaire analysis will also be stored on this server (See the section entitled Data Storage).
- *Sampling:* The sampling frame for the follow-up questionnaire includes all visitors who completed an interview at the museum and who voluntarily supplied the ILI project team with their email address. For those visitors younger than 18 years old, we will email the invitation to

the questionnaire to the adult who is supervising the minor at the museum. For those 18 and older, the email with the invitation to the questionnaire will go directly to the respondent.

- *Consent and Recruitment:* Using the contact information (i.e. email address) collected at the end of the on-site interview, visitors will be sent an email invitation to the online questionnaire six weeks after their museum visit. The email invitation will be generated by Qualtrics from ILI research staff containing an invitation to the questionnaire as well as the questionnaire link (a template of this invitation email is included in Appendix C). Within this invitation, visitors will be reminded that they gave their email address while at the museum and the purpose of the questionnaire. Once visitors have clicked on the questionnaire link they will be taken to an introduction page which will include the consent information included at the beginning of the *Twins Follow-up Questionnaire* instrument. To consent to participate in the questionnaire and to begin the questionnaire, visitors actively select the "Start Questionnaire" button. A link between the email address to which the invitation was sent and the data that is completed by the individual with that email address will allow researchers to connect the follow-up data with the data collected at MOS.
- *Reminders:* ILI researchers will use the capabilities of Qualtrics to send a follow-up reminder email only to those visitors who have not completed the questionnaire. Reminders will be sent one week after the original invitation email (a template for the follow-up email is included in Appendix C).
- *Incentive:* At the end of the online questionnaire, visitors will be asked to provide an email address which will be used to deliver the incentive, a $10 gift certificate to Amazon.com.

Coach Mike at Robot Park

The summative evaluation will include measures of visitor interaction and impacts at Robot Park, where Coach Mike is included as part of the current exhibit. The evaluation of the Coach Mike at Robot Park will include visitor observation, visitor interviews, and follow-up online questionnaire. The instruments to be used and the consent, sampling, recruitment, and incentive procedures for each of these methods is detailed below.

A. Visitor Observations

- *Instrument:* Observations of visitors to Robot Park will be recorded using the *Coach Mike Observation* Instrument. This instrument will be used to record visitor interaction with the exhibit's tangible interface, interaction between visitors and visitors and MOS staff, easily observed behaviors relating to Coach Mike, and any usability issues. See Appendix B for the guide to how visitor behaviors will be recorded on the *Coach Mike Observation* Instrument.
- *Passive Consent:* While ILI staff members are present in CCP and conducting observations, a sign notifying visitors of the observation will be posted near Robot Park. The sign will read: "To help us better serve our visitors, this area is being observed ending at [time]. With any questions or concerns, please contact Christine Reich at 617-589-0302."
- *Sampling:* The sampling frame for the observations includes all visitors who appear to be at least 7 years old and who interact with the Robot Park exhibit. For the purposes of this observation, "interaction" with the exhibit is defined as creating (or attempting to create) a program using the tangible interface. In order to select the "target" visitor who will be the focus of an observation, ILI staff will use the following method:
    - When the data collector is ready to begin an observation (i.e. immediately after a break or after completing an interview), they will station themselves within 5 to

10 feet of Robot Park with a direct line of sight to the tangible interface work space.
    - The next visitor who appears to be at least 7 years old and who 1) plants both feet in front of the exhibit or 2) sits on the stool will be the target for the observation and timing will start. If the target does not interact with the exhibit (as defined above), the data collector will observe the next available visitor who meets the above criteria, restarting the timing.
    - The observation will continue as long as the target visitor is 1) interacting with Robot Park or 2) watching another visitor interact with Robot Park. Once the visitor either 1) moves to exit CCP or 2) begins to engage with another exhibit, the data collector will approach the target and request an interview.
- *Recruitment:* No recruitment method will be used with visitors before beginning the observations.
- *Incentive:* There are no incentives given to visitors in exchange for their participation in the observations.

B. Visitor Interviews

- *Instrument:* Interviews with visitors to Robot Park will be guided by the *Coach Mike Interview* Instrument. This instrument will be used to ask visitors open-ended and rating scale question about 1) their interactions with Coach Mike and Robot Park, 2) changes in their knowledge, attitudes, awareness, and engagement as a result of their interactions, 3) basic demographic and psychographic information, and 4) an email address for the purposes of contacting visitors for a follow-up web questionnaire. The interview will take approximately 5 minutes to complete.
- *Sampling:* The sampling frame for the interviews includes all visitors who were targets of an observation, with the goal to create a matched observation of and interview with the same target visitor. After the observation of a target visitor ends (as defined in the Visitor Observation section above), the ILI staff member will approach the target visitor for an interview. While other members of the target visitor's group may participate in the interview, the data collector will ensure the target answers all questions and supplies their demographic and psychographic information.
- *Consent and Recruitment:* After the observation of a target visitor ends, the ILI staff member will approach the target visitor for an interview. The script included at the beginning of the *Coach Mike Interview* instrument will serve as both the consent and recruitment language for visitors. Visitors who agree to speak with the interviewer have consented to participate in the interview. Those visitors who appear to be under 16 years of age will be asked for their assent to participate in the interview *after* the adult who is supervising them gives their consent to the interview. In the event that a target visitor declines to participate in an interview (or have their child under 16 participate in an interview), the visitor will be thanked by the ILI staff member and allowed to continue their visit; the observational data from visitors who decline to participate in an interview will be maintained by ILI project staff and analyzed 1) as a point of comparison to those visitors who did consent to the interview and 2) to increase the overall sample size of the observations.
- *Incentive:* In thanks for completing the interview, visitors will be offered a small thank-you gift with a value of less than $1, such as a pencil.

C. Follow-up Questionnaire

- *Instrument:* A follow-up online questionnaire with target visitors will be guided by an online version of the *Coach Mike Follow-up Questionnaire* Instrument. This instrument will be used to ask visitors open-ended and rating scale question about their knowledge, attitudes, awareness, and engagement outcomes as a result of their interactions with Coach Mike and Robot Park at MOS and after they left the museum. The questionnaire will take approximately 5 minutes to complete. The online questionnaire will be hosted by Qualtrics, a program with which ILI researchers have extensive experience. Qualtrics allows the ILI project staff to create and maintain the online questionnaire in a password-protected environment accessible only to the ILI project staff. This maintains the safety of the data while it is within the Qualtrics system. Once all questionnaire data has been collected, the data will be downloaded to ILI's secure server. Questionnaire analysis will also be stored on this server (See the section entitled Data Storage).
- *Sampling:* The sampling frame for the follow-up questionnaire includes all visitors who completed an interview at the museum and who voluntarily supplied the ILI project team with their email address. For those visitors younger than 18 years old, we will email the invitation to the questionnaire to the adult who is supervising the minor at the museum. For those 18 and older, the email with the invitation to the questionnaire will go directly to the respondent.
- *Consent and Recruitment:* Using the contact information (i.e. email address) collected at the end of the on-site interview, visitors will be sent an email invitation to the online questionnaire six weeks after their museum visit. The email invitation will be generated by Qualtrics from ILI research staff containing an invitation to the questionnaire as well as the questionnaire link (a template of this invitation email is included in Appendix C). Within this invitation, visitors will be reminded that they gave their email address while at the museum and the purpose of the questionnaire. Once visitors have clicked on the questionnaire link they will be taken to an introduction page which will include the consent information included at the beginning of the *Coach Mike Follow-up Questionnaire* instrument. To consent to participate in the questionnaire and to begin the questionnaire, visitors actively select the "Start Questionnaire" button. A link between the email address to which the invitation was sent and the data that is completed by the individual with that email address will allow researchers to connect the follow-up data with the data collected at MOS.
- *Reminders:* ILI researchers will use the capabilities of Qualtrics to send a follow-up reminder email only to those visitors who have not completed the questionnaire. Reminders will be sent one week after the original invitation email (a template for the follow-up email is included in Appendix C).
- *Incentive:* At the end of the online questionnaire, visitors will be asked to provide an email address which will be used to deliver the incentive, a $10 gift certificate to Amazon.com.

Risks and Benefits to the Visitors

There are no known physical, social, knowledge-related, or financial risks associated with participating in the evaluation study.

Potential risks to visitors as a result of the evaluation study:

- *Privacy risks*: The loss of privacy is a possible, though unlikely, risk for visitors who volunteer their email addresses for the purposes of the follow-up web questionnaire. ILI will maintain the email addresses in compliance with the Data Storage procedures outlined in that section below. Furthermore, ILI will use the email addresses for no other purpose except to contact visitors with the invitation to the follow-up web questionnaire, reminders to complete the web

questionnaire, and to send the $10 gift certificate to those who complete the questionnaire. The email addresses will not be linked to other personally identifiable information and will not be accessible to those outside the ILI project team.

Possible benefits to visitors as a result of the evaluation study:

- *Personal benefits*: The visitors participating in the evaluation may benefit from the study process by having the opportunity to share their views on the museum experiences being evaluated. ILI researchers have found that visitors generally like being asked to give their opinion about museum programs or exhibitions and enjoy having the opportunity to reflect on their experience and hear the opinions of others in their group (especially children).

Possible benefits to society as a result of the evaluation study:

- *Benefits for the field of Informal Science Education*: As a result of this evaluation study, the field of informal science education will have a better understanding of how visitors to a science museum interact with and perceive the usefulness of virtual humans. This may help other institutions decide whether to use virtual humans in a museum setting and how virtual humans may best be used with visitors.

Reporting

In the draft and final versions of the report, at no time will any personally identifiable information be connected to visitors' responses.

Data Storage

Physical records will be kept in a secured filing cabinet in the research offices of the Institute for Learning Innovation in Edgewater, Maryland. This cabinet will be locked and the key kept by a member of the ILI staff who is on the project team. Electronic versions of records or databases consolidating the data will be maintained on a secure server at the Institute for Learning Innovation and password-protected.

Violations, Adverse Events, or Changes to the Protocol

Violations to this protocol will be reported to the IRB as stipulated by their policies. Adverse events occurring during this course of the study will also be reported to IRC as stipulated by their policies. Any needed changes to the protocol will also be submitted to the IRB before they are acted upon.

Protocol Appendix A: Twins Observation Guide

This appendix serves as a guide for the on-site observations of the Twins and The Science Behind exhibits using the *Twins Observation* Instrument. The ILI project team who are collecting the observational data will use this as a training document and in the event questions related to the observation arrise.

Twins Interactions: A chart format has been developed to allow for easy tracking and recording of the interactions between visitors, staff, and the Twins. The goal of this format is to accurately capture the

questions asked by visitors or staff, the response of the Twins, and the overall flow and progression of the conversation. The chart is to be used for recording conversations with the Twins NOT conversations between visitors or between visitors and staff. The data collector will begin using the chart when the target visitor starts their interaction with the Twins (as defined in the Visitor Observation section of the protocol). Each area of the chart is explained below:

- Who: The data collector will mark who asks the first question to the Twins. V= the target visitor of the observation. S= a CCP staff member. O= other visitor, not the target.
- Utterances: The central area of the chart will be used to record the questions or statements the target visitor, other visitors, or staff members address to the Twins. Addressing a question to the Twins is defined as a person pushing (or attempting to push—i.e. misuse of the button is counted as an attempt) the talk button and asking the Twins a question/making a statement to the Twins. Each question the target visitor, a staff member, or another visitor addresses to the Twins will be recorded and the appropriate response category marked (explained below). The data collector will write the question/statement exactly as it is asked. If a question is repeated exactly, the data collector may use a ditto mark (i.e. ") in the space below the repeated question to indicate that it was repeated exactly. A variation in how a question was asked should be written word for word and not marked as a repeated question (i.e. "What are your names?" is different from "Tell me your names.")
    - Intro/Greetings: For introductions and saying hello to the Twins. Includes "Hello", "Hi", "What are your names?"
    - Personal: For questions/statements address to the Twins about themselves that go beyond introductions. This includes "Who are you named after?", "What do you like to do for fun?", "Are you twins?", "What's your favorite color?", "Do you have a boyfriend?", "Why are you wearing a red coat?", "Do you like the Red Sox?"
    - Tech/Computer: For questions/statements addressed to the Twins about technology, computers, and objects or activities in CCP (including robots, artificial intelligence, virtual humans, and cell phones). Includes questions like "Are you a computer?", "What is artificial intelligence?", "Tell me about cell phones.", "Where can I go to learn about robots?"
    - General Museum/Other: For questions/statements addressed to the Twins about the museum in general (excluding CCP) or other topics not covered by the other categories. Includes "Where are the bathrooms?", "How do I find the dinosaurs?", "What time is the lightning show?", "What is there to see in Boston?", "Will the Red Sox beat the Orioles?"
- Twins: These columns are to record the responses of the twins to the question/statement asked of them.
    - Response: The data collector will check off a box to indicate to what degree the response of the Twins matched the question/statement addressed to them. The data collector will use their best judgment to categorize the responses.
        - Aprp: Appropriate response to the question/statement. Mark this if the answer is a reasonable answer to what was addressed to the Twins. Reasonable means on-topic or like a response you might expect from a person. Example of a dialogue with an appropriate response: "What are your names?" "I'm Ada. And I'm Grace."
        - Inaprp: Inappropriate response to the question/statement. Mark this if the answer is an unreasonable answer to what was addressed to the Twins. Unreasonable means off-topic or unrelated to what was asked. Example of a

dialogue with an inappropriate response: “What are your names?” “A tangible interface is …”

- DK: Don’t know response. Any response of the Twins that indicates that they do not understand the question. Causes of a DK response include user error, an inaccuracy in the voice recognition, or they simply aren’t programmed to have the answer. Examples of a DK response include asking the visitor to repeat what they said, saying they didn’t hear the visitor, or other less obvious non-answers.

- Suggests? Sometimes the Twins suggest a visitor does something. For example, visit The Science Behind exhibit, go to another exhibit like Robot Park, or “You could ask us about cell phones.” Mark these suggestions in the Suggest column. If no suggestion is made, leave the column blank. It is ok to use shorthand here as long as it is easily understood by the data collector after the fact (i.e. *SB*= they suggested the visitor go to Science Behind; *ask cell*= they suggested the visitor ask about cell phones).

Time spent: Time spent at the two exhibits will be measured in minutes and seconds with either a stopwatch or a stop watch-like app on a mobile phone. Timing and observation will start when the next visitor who appears to be at least 7 years old 1) plants both feet in front of the podium on which the microphone is mounted or 2) sits on the stool will be the target for the observation and timing will start.

Timing and observation will continue as long as the target visitor is 1) interacting with the Twins, 2) watching another visitor interact with the Twins, or 3) engaging with The Science Behind. The timing is continuous between the Twins and The Science Behind; keep timing as a visitor moves between the two spaces. If the target visitor leaves and then returns to the Twins or The Science Behind, timing will be restarted and both times added together for a total time. Once the visitor either 1) moves to exit CCP or 2) begins to engage with another exhibit, the observation ends.

Observable Behaviors: Data collectors will record easily observable behaviors occuring at the Twins exhibit at the bottom of the *Twins Observation* instrument. These include:

- Asks questions suggested on label: Mark if the visitor asks a question that is on the label. This can be filled out after the interaction if the data collector needs to verify the question on the label.
- Visits Science Behind *(over)*: Mark if a visitor goes to The Science Behind. IF THIS IS CHECKED, the data collector must complete the Science Behind Interactions section on the instrument.
- Repeats a question exactly: Mark if the exact same question is asked by a visitor or staff member.
- Other: Fill in any other easily observable behaviors that occur at the Twins. If more space is needed, please indicate (back) in the blank, and turn the instrument over and record the comments/interactions in the space provided. This could include a more in-depth description of signification social interactions as well as interactions with the Twins.

Usability: Usability issues include hardware or user errors that occur while the target visitor is at the Twins. Each of the following will be checked by the data collector should they occur.

- Button Misuse: Includes the visitor not holding down the button when talking, pushing the button too many times in a row, pushing the button in a way that seems to be interfering with the Twins ability to respond.

- Speech Issues (volume/accent): If there is a reason to suspect that inappropriate or DK responses are due to a speech issue, the data collector should mark this box. Not every instance of an inappropriate or DK response is due to a speech issue. Only the most severe issues (a very heavy accent, speaking in a language other than English, yelling or whispering into the microphone) should be marked. The data collector should use their own judgment when marking this box.
- Frozen computer: If the Twins stop responding to a visitor, seem to freeze up, or need to be rebooted by a staff member, this box should be marked.
- Other: Other usability issues should be explained here.

Science Behind Interactions: Data collectors will record easily observable behaviors occuring at the Science Behind exhibit on the back of the *Twins Observation* instrument. This includes social interactions and a a description of interactions at the exhibit (See below). If this section is completed only if the visitor actually goes to the Science Behind. Please be sure to also check the "Visits Science Behind" box on the front of the instrument.

Description of Interaction with Science Behind *(record here):* We are not taking running, ethnographic style notes about the interactions at The Science Behind. However, we do want to capture significant interactions and behaviors at the exhibit. The data collector should note if they can tell what the visitor is looking at either by their sight line or other behaviors such as pointing. Please use brackets [] around anything that is not a direct quote and quotation marks " " around direct quotes from a visitor or staff member. For example when piloting we observed the following: [boy and target run to SB, boy stays to look at voice recognition output; target goes back to Twins, asks question]. Social interactions could be captured in this area, but MUST be noted using the "Social Interactions" checkboxes.

Social Interactions: Target to Visitor: Defined as interactions (typically verbal) between the target visitor and any other visitor whether they are in the target's group or not. Target to visitor social interactions may occur at the Twins or Science Behind; social interactions about/near the Twins should be noted on the front of the instrument, and social interactions about/near Science Behind on the back of the instrument. Each interaction will be checked by the data collector as it occurs; the data collector must choose the appropriate category. If the data collector is not sure what category an interaction falls under, the interaction should be included in the "Other" option.

- Conceptual: This category is for conceptual talk about the purpose of the exhibit or the concepts the exhibit is trying to teach. The talk may or may not be sophisticated or not; for example "Here you get to talk to these girls" is less sophisticated than "Here you can talk to the virtual humans Ada and Grace" but both are examples of conceptual speech. Also includes general questions such as "What is this about?" or "What do you do here?"
- Operational: This category is for talk about how to operate the exhibit or what specifically to ask the Twins. Utterances may be *instructions* ("Push the button", "Sit on the stool", "Lean closer to the microphone", "Ask them about cell phones"), *suggestions* ("You could sit on the stool", "Try pushing the button", "Ask them something you would ask a friend") or a *question* ("What do you think this button does?", "What will happen when I talk to them?", "What do they know about?"). Also includes statements such as "I don't know what to ask" and "Tell me what to ask."
- Positive Affect: Any positive talk or utterance. Includes "Cool!", "They understood!", and laughing.

- Negative Affect: Any negative talk or utterance. Includes “It’s not working”, “They are dumb”, “Boo!”
- Other: Any interaction that does not clearly fit into one of the categories above should be explained in this category. This includes instances where the visitors cannot be clearly heard but can be seen to be talking and instances where the visitors are speaking a language that the data collector does not know. Please fill in the blank with the word, phase, or utterance OR a bracketed summary of the issue (i.e. [Speaking German], [Too quiet for me to hear]).

Social Interactions: Target to Staff: Defined as verbal interactions between the target visitor and a CCP staff member. Each interaction will be checked by the data collector as it occurs; the data collector must choose the appropriate category. Target to staff social interactions may occur at the Twins or Science Behind; social interactions about/near the Twins should be noted on the front of the instrument, and social interactions about/near Science Behind on the back of the instrument. Note that interaction between the visitor and a staff member may be at a higher level than that between two visitors. However the same general categories still apply. If the data collector is not sure what category an interaction falls under, the interaction should be included in the “Other” option.

- Conceptual: This category is for conceptual talk about the purpose of the exhibit or the concepts the exhibit is trying to teach. The talk may or may not be sophisticated or not; for example “Here you get to talk to these girls” is less sophisticated than “Here you can talk to the virtual humans Ada and Grace” but both are examples of conceptual speech. Also includes general questions such as “What is this about?” or “What do you do here?” A staff member might also explain what the Twins are, how they were made, or their programming and the technology they use OR more about computers or other technology related to the exhibit.
- Operational: This category is for talk about how to operate the exhibit or what specifically to ask the Twins. Utterances may be *instructions* (“Push the button”, “Sit on the stool”, “Lean closer to the microphone”, “Ask them about cell phones”), *suggestions* (“You could sit on the stool”, “Try pushing the button”, “Ask them something you would ask a friend”) or a *question* (“What do you think this button does?”, “What will happen when I talk to them?”, “What do they know about?”). Also includes statements such as “I don’t know what to ask” and “Tell me what to ask.” A staff member also might pose step-by-step questions or instructions to help a visitor understand usability issues (i.e. such as waiting until the Twins go through their Listening-Thinking-Responding routine before asking another question); this counts as operational. However, if the staff member then takes that explanation further (i.e. “A computer needs to take one input/instruction at a time”) that may count as conceptual as well; the data collector would then mark both operational and conceptual.
- Positive Affect: Any positive talk or utterance. Includes “Cool!”, “They understood!”, “You did it!” and laughing.
- Negative Affect: Any negative talk or utterance. Includes “It’s not working”, “They are dumb”, “Boo!”
- Other: Any interaction that does not clearly fit into one of the categories above should be explained in this category. Please fill in the blank with the word, phase, or utterance OR a bracketed summary of the issue (i.e. [Speaking German], [Too quiet for me to hear]).

Protocol Appendix B: Coach Mike Observation Guide

This appendix serves as a guide for the on-site observations of Coach Mike at Robot Park using the *Coach Mike Observation* Instrument. The ILI project team who are collecting the observational data will use this as a training document and in the event questions related to the observation arrise.

Time spent: Time spent at the interaction will be measured in minutes and seconds with either a stopwatch or a stop watch-like app on a mobile phone. Timing and observation will start when the next visitor who appears to be at least 7 years old 1) plants both feet in front of the exhibit or 2) sits on the stool. Timing and observation will continue as long as the target visitor is 1) interacting with Robot Park or 2) watching another visitor interact with Robot Park. If the target visitor leaves and then returns to Robot Park, timing will be restarted and both times added together for a total time. Once the visitor either 1) moves to exit CCP or 2) begins to engage with another exhibit, the observation ends.

Observed Behaviors: A set list of observable behaviors is included on the observation instrument. The data collector will check each behavior the target visitor is seen to engage in. A description of each observable behavior is included below

- Looks at the Screen *(When Mike is OFF)*: The target visitor looks at the monitor screen mounted on the back wall of Robot Park. This behavior applies only when Coach Mike is off (or not engaged). If the visitor looks at the screen when Coach Mike is on, the data collector will check the "Looks when talking" behavior (See below under the section Interacts with Coach Mike).
- Physically moves the robot: The target visitor (or someone in their group or a staff member) picks up and moves the robot during the observation.
- Uses Block Tester: The target visitor places at least one block in the "Block Tester" area of the exhibit.
- Other: Any other generally observable behavior that is not captured elsewhere on the sheet AND that the data collector feels is important will be noted here.

Challenges Attempted and Completed: Every challenge the target visitor tries to complete will receive a checkmark by the data collector under the "Attempted" column. Indications of an attempt include 1) talking to another visitor or a staff member about how to complete the challenge or 2) assembling (or reassembling) a series of blocks in a way that the data collector judges is an attempt to meet the challenge. Every challenge that is successfully completed by the target visitor will receive a checkmark by the data collector in the "Completed" column (See below for an explanation of each challenge).

- Lights up sign: A sign that says "Target" is under the monitor along the back wall of Robot Park. A visitor successfully completes this challenge when they move the robot to the Target sign and the overhead "Robot Park" sign lights up and moves.
- Turn around: Turning around is a challenge issued by Coach Mike; he congratulates the visitor when they have successfully completed the challenge. It involves moving the robot in a tight circle, but without the use of the "Spin" block. Solutions include 3 or 4 right blocks or 3 or 4 left blocks in a row.
- Make a square: Make a square is a challenge issued by Coach Mike; he congratulates the visitor when they have successfully completed the challenge. It involves moving the robot in a square. Solutions include right-forward-right-forward-right-forward-right-forward or a similar combination using lefts or reverses.

Interacts with Coach Mike: The data collector will look for and record behaviors that indicate that the target visitor has noticed and is paying attention to Coach Mike. Behaviors in this category will be marked only then Coach Mike is engaged.

- Mike button, # of times: Checked if the target visitor pushes the Coach Mike help/hint button; the number of times the button is push will also be recorded as a series of "tic" marks.
- Looks when talking: The target visitor appears to look up at the monitor screen when Mike is talking.
- Mentions to another visitor/staff member: The target visitor makes some comment that refers to Coach Mike in the judgment of the data collector.
- Tries to follow his direction/suggestion: Coach Mike often makes suggestions to visitors. This includes relatively simple instructions (i.e. "Find the start block. Place it in the block tester." "Connect the start block to the forward block."), issuing challenges (i.e. "See if you can get the robot to turn around."), or hints to solve a challenge. If the target visitor appears to follow a suggestion immediately after Coach Mike has made the suggestion, this item will be checked by the data collector.
- Other: Any other behavior that indicates the target visitor has noticed or is paying attention to Coach Mike will be indicated here.

Programs with Run pressed *(enter # of blocks)*: The data collector will enter a number into the chart on the observation sheet every time the target visitor presses the "Run" button. When the Run button is pushed, a clicking sound is made (like a camera taking a picture) and the program that the robot will follow appears on the monitor. The count will include the start block and any other blocks that appear on the screen as an active part of the program. For example, just the start block would count as 1, the start and another block as 2, etc. The following instances also will be counted by the data collector:

- Unintentional pushes of the Run button by the target visitor.
- Pushing Run by someone other than the target when the target is actively programming the robot.
- Pushing Run for a program that they did not create (i.e. one that existed in the work area when the visitor started their interaction).
- If part of the camera's shot is blocked (by a visitor's hand, head, etc), only blocks that appear on the screen as an active part of the program will be counted.

Do not count programs created by another member of the target visitor's group.

Usability Issues: Usability issues include hardware or user errors that occur while the target visitor is at Robot Park. Each of the following will be checked by the data collector should they occur. Camera: The visitor is in the way of the camera when the button is pushed. The only way to judge this is by looking at the image on the screen to see if something is blocking part of the program.

- Mat: The visitor is 1) not fully on the mat or 2) does not weigh enough to trigger the mat. This condition will be judged based on the responses of Coach Mike ONLY. Coach Mike will repeatedly greet and say good-bye to visitors who have not actually left when there is an issue with the mat.
- Frozen computer/crash: The computer may have a prolonged rotating image of a flower when this occurs and may have to be rebooted by the CCP staff.
- Other: Any other usability issues that occur while the target visitor is at the exhibit. Please note that the visitor's trial and error attempts to figure out how the exhibit works is not a usability issue for this study.

Social Interactions: Target to Visitor: Defined as interactions (typically verbal) between the target visitor and any other visitor whether they are in the target's group or not. Each interaction will be checked by the data collector as it occurs; the data collector must choose the appropriate category. If the data

collector is not sure what category an interaction falls under, the interaction should be included in the "Other" option.

- Conceptual: This category is for conceptual talk about the purpose of the exhibit or the concepts the exhibit is trying to teach. The talk may or may not be sophisticated or not; for example "Here you try to get the robot to move" is less sophisticated than "Here you program the robot to make it move" but both are examples of conceptual speech. Also includes general questions such as "What is this about?" or "What do you do here?"
- Operational: This category is for talk about how to operate the exhibit. Utterances may be *instructions* ("Use the forward block", "First put it in the block tester"), *suggestions* ("Maybe try the forward block", "If you want to try the block tester") or a *question* ("What do you think the forward block will do?", "What do you think the block tester does?"). Also includes questions such as "What do you do first?" or "How do I get started?"
- Goal setting: This category is for talk about completing the challenge or setting another goal. Examples include "Let's try to hit the sign", "Can we make it go in a square?", "I'm going to make it go to that corner." Note that individual steps the visitor talks about to reach the goal most likely will fall under the Operational category.
- Positive Affect: Any positive talk or utterance. Includes "You did it!", "Cool!", "Yes!", and laughing.
- Negative Affect: Any negative talk or utterance. Includes "This is hard", "I can't do it", "Boo!"
  Other: Any interaction that does not clearly fit into one of the categories above should be explained in this category. This includes instances where the visitors cannot be clearly heard but can be seen to be talking and instances where the visitors are speaking a language that the data collector does not know. Please fill in the blank with the word, phase, or utterance OR a bracketed summary of the issue (i.e. [Speaking German], [Too quiet for me to hear]).

Social Interactions: Target to Staff: Defined as verbal interactions between the target visitor and a CCP staff member. Each interaction will be checked by the data collector as it occurs; the data collector must choose the appropriate category. Note that interaction between the visitor and a staff member may be at a higher level than that between two visitors. However the same general categories still apply. If the data collector is not sure what category an interaction falls under, the interaction should be included in the "Other" option.

- Conceptual: This category is for conceptual talk about the purpose of the exhibit or the concepts the exhibit is trying to teach. The talk may or may not be sophisticated or not; for example "Here you try to get the robot to move" is less sophisticated than "Here you program the robot to make it move" but both are examples of conceptual speech. Also includes general questions such as "What is this about?" or "What do you do here?" A staff member might also give an explanation what programming is, how it works, why it is needed; these count as conceptual.
- Operational: This category is for talk about how to operate the exhibit. Utterances may be *instructions* ("Use the forward block", "First put it in the block tester"), *suggestions* ("Maybe try the forward block", "If you want to try the block tester") or a *question* ("What do you think the forward block will do?", "What do you think the block tester does?"). Also includes questions such as "What do you do first?" or "How do I get started?" A staff member also might pose step-by-step questions or instructions to help a visitor complete a goal; this counts as operational.
- Goal setting: This category is for talk about completing the challenge or setting another goal. Examples include "Let's try to hit the sign", "Can we make it go in a square?", "I'm going to make

it go to that corner." Note that individual steps the visitor talks about to reach the goal most likely will fall under the Operational category.

- Positive Affect: Any positive talk or utterance. Includes "You did it!", "Cool!", "Yes!", and laughing.
- Negative Affect: Any negative talk or utterance. Includes "This is hard", "I can't do it", "Boo!"
- Other: Any interaction that does not clearly fit into one of the categories above should be explained in this category. Please fill in the blank with the exact word, phase, or utterance in quotes OR a bracketed summary of the issue (i.e. [Speaking German], [Too quiet for me to hear]).

Significant Interactions *(record here):* We are not taking running, ethnographic style notes about the interactions. However, if something note-worthy or out of the ordinary occurs, please note it in this section. For example, during piloting a father was heard telling his son *not* to use the "Hint" button; so Mike was never triggered for this reason. Another example, might include notes on how the target moves from being the main programmer to more of an observer or coach (i.e. [Target moved so sister could sit, stayed to help her out.]) or notes on interactions with staff (i.e. [Got staff member's attention because they couldn't figure out how to make it work.]) Please use brackets [] around anything that is not a direct quote and quotation marks " " around direct quotes from a visitor or staff member.

Protocol Appendix C: Follow-up Questionnaire Invitation and Reminder Emails

Initial email invitation for the follow-up online questionnaire: Adult visitor
To be sent to visitors who volunteered their email addresses and agreed to participate in the follow-up questionnaire when they were interviewed at the museum. To be sent six weeks after their visit.

To: [visitor's email address]
From: Susan Foutz (foutz@ilinet.org)
Subject line: Following-up on your visit to the Museum of Science, Boston

Email text:
Dear Museum of Science visitor,

A few weeks ago you were kind enough to participate in an interview at the Museum of Science and agreed to participate in a follow-up questionnaire. We are conducting the follow-up questionnaire to learn more about how your visit to the museum and interactions with the exhibits impacted you. We appreciate your willingness to participate in the questionnaire.

Please click on this link to take the follow-up questionnaire [Questionnaire Link].

This questionnaire takes about [5 minutes for Robot Park/10 minutes for the Twins] to complete. Everyone who completes the questionnaire will receive a $10 gift certificate for Amazon.com. At the end of the questionnaire, you'll be prompted to enter an email address at which to receive the gift certificate. Please complete the questionnaire by [enter date here. The closing date will be 7-10 days after initial send, the 2nd Friday after the initial invite is preferred.]

If you have any questions about the questionnaire or anything related to this email, please feel free to contact me.

Thanks so much for your help,

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
410-956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

Reminder email for the follow-up online questionnaire: Adult visitor

To be sent to visitors who 1) volunteered their email addresses and agreed to participate in the follow-up questionnaire when they were interviewed at the museum and 2) did not complete the questionnaire after the initial invitation. To be sent three days after the closing date named in the invitation email.

To: [visitor's email address]
From: Susan Foutz (foutz@ilinet.org)
Subject line: Following-up on your visit to the Museum of Science, Boston

Email text:
Dear Museum of Science visitor,

Last week, we emailed you with an invitation to complete a short web questionnaire about your visit to the Museum of Science. You received the invitation as a result of generously agreeing to participate in the questionnaire when you visited the museum six weeks ago. We realize you are very busy and wanted to give you another chance to fill out the questionnaire in case you ran out of time.

Please take 5 minutes to fill out the questionnaire. Everyone who completes the questionnaire will receive a $10 gift certificate for Amazon.com. At the end of the questionnaire, you'll be prompted to enter an email address at which to receive the gift certificate.

This link will take you directly to the questionnaire [Questionnaire Link]. If possible, we would love to hear from you by [extension date- that Friday after the reminder email is sent].

If you have any questions about the questionnaire or anything related to this email, please feel free to contact me.

Thanks for your participation,

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
410-956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

Initial email invitation for the follow-up online questionnaire: Child visitor
To be sent to Adult visitors 1) whose child participated in the on-site interview and 2) who volunteered their email addresses and agreed to participate in the follow-up questionnaire when they were interviewed at the museum. To be sent six weeks after their visit.

To: [visitor's email address]

From: Susan Foutz (foutz@ilinet.org)
Subject line: Following-up on your visit to the Museum of Science, Boston

Email text:
Dear Museum of Science visitor,

A few weeks ago your child participated in an interview at the Museum of Science and you agreed to participate in a follow-up questionnaire with them. We are conducting the follow-up questionnaire to learn more about how the visit to the museum and interactions with the exhibits impacted your child. We appreciate your willingness to participate in the questionnaire with your child.

Please click on this link to take the follow-up questionnaire [Questionnaire Link].

This questionnaire takes about [5 minutes for Robot Park/10 minutes for the Twins] to complete. Everyone who completes the questionnaire will receive a $10 gift certificate for Amazon.com. At the end of the questionnaire, you'll be prompted to enter an email address at which to receive the gift certificate. Please complete the questionnaire by [enter date here. The closing date will be 7-10 days after initial send, the 2nd Friday after the initial invite is preferred.]

If you have any questions about the questionnaire or anything related to this email, please feel free to contact me.

Thanks so much for your help,

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
410-956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

Reminder email for the follow-up online questionnaire: Child visitor

To be sent to Adult visitors 1) whose child participated in the on-site interview, 2) who volunteered their email addresses and agreed to participate in the follow-up questionnaire when they were interviewed at the museum, and 3) did not complete the questionnaire after the initial invitation. To be sent three days after the closing date named in the invitation email.

To: [visitor's email address]
From: Susan Foutz (foutz@ilinet.org)
Subject line: Following-up on your visit to the Museum of Science, Boston

Email text:
Dear Museum of Science visitor,

Last week, we emailed you with an invitation to complete a short web questionnaire about your child's experience at the Museum of Science. You received the invitation as a result of generously agreeing to participate in the questionnaire when you and your child visited the museum six weeks ago. We realize you are very busy and wanted to give you another chance to fill out the questionnaire in case you ran out of time.

Please take 5 minutes to fill out the questionnaire. Everyone who completes the questionnaire will receive a $10 gift certificate for Amazon.com. At the end of the questionnaire, you'll be prompted to enter an email address at which to receive the gift certificate.

Please complete the questionnaire with your child. This link will take you directly to the questionnaire [Questionnaire Link]. If possible, we would love to hear from you by [extension date- that Friday after the reminder email is sent].

If you have any questions about the questionnaire or anything related to this email, please feel free to contact me.

Thanks for your participation,

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
410-956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

Appendix 2 **Instruments: The Twins**

Twins Observation
ICT-MOS Summative Evaluation

Date: ______ Time: ________ Data Collector: ________ Interview ID #:________
Condition: ☐ Direct ☐ Blended—Handler: ______________________

Twins Interactions

Who	Utterance	Twins	
		Response	Suggests?
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	
☐ V ☐ S ☐ O	☐ Intro/Greeting ☐ Personal ☐ Tech/Computer ☐ Gen. Museum/Other	☐ Aprp ☐ Inaprp ☐ DK	

Target to Visitors	Target to Staff
☐ Conceptual	☐ Conceptual
☐ Operational	☐ Operational
☐ Positive affective	☐ Positive affective
☐ Negative affective	☐ Negative affective
☐ Other: _______________	☐ Other: ______________

Time spent:___________(min : sec)
☐ Asks questions suggested on label ☐ Visits Science Behind *(over)*
☐ Repeats a question exactly ☐ Other: ______________________

Usability
☐ Button Misuse ☐ Speech Issues (volume/accent) ☐ Frozen computer
☐ Other: ____________

Twins Observation
ICT-MOS Summative Evaluation

Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____
Condition: ☐ Direct ☐ Blended—Handler: _____

Science Behind Interactions

Target to Visitors	**Target to Staff**
☐ Conceptual	☐ Conceptual
☐ Operational	☐ Operational
☐ Positive affective	☐ Positive affective
☐ Negative affective	☐ Negative affective
☐ Other	☐ Other

Description of Interaction with Science Behind *(Record here. Note conversations that happen while at Science Behind, whether they read the labels, read the screen, search for their questions and the twins response, number of times they move between Science Behind and the Twins)*

Observed Demographics *(complete only in the absence of interview demographics)*

Target's gender:

- ☐ Female
- ☐ Male

Target's age range

- ☐ Under 18
- ☐ Over 18

Children in the Target's Group

- ☐ Yes, saw kids
- ☐ No kids, only saw adults

Other Signification Interactions at the Twins *(record here)*:

Twins Interview Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____
ICT-MOS Summative Evaluation Condition: ☐ Direct ☐ Blended

Hi, the Museum is trying to get feedback about the exhibit you just used, and I was wondering if you would be willing to answer a few questions? It will take 5-10 minutes. FOR PARENTS OF CHILDREN: *Hi, the Museum is trying to get feedback about the exhibit your child just used, and I was wondering if I could interview your daughter/son? It will take 5-10 minutes.* IF YES: *Before we start, I want to let you both know that you don't have to answer the questions if you don't want to and that you can end the interview at any time. There are also no right or wrong answers to my questions. Any thoughts you share will help us improve our exhibits.*

1. What did you find most interesting about this exhibit?

2. How would you describe the Twins, Ada and Grace, to someone else?

3. The computer scientists who designed Ada and Grace say they are examples of a type of technology called a "virtual human". a. In what ways do you think Ada and Grace are the same as a real person?

   b. In what ways are Ada and Grace different from a real person?

4. What technologies do you think are needed to make Ada and Grace work?

   a. *(if "I don't know")* Do you think it needs a computer to work? ☐ Yes ☐ No ☐ Unsure
   b. *(If Yes)* How do you know it needs computers?

5. What would you say was the main idea of the exhibit? (Probe: What was the museum hoping visitors would learn?)

6. Did you learn anything new about technology or computers from interacting with the exhibit?
   ☐ Yes ☐ No ☐ Unsure
   a. *(If Yes)* What did you learn?

7. Did interacting with Ada and Grace help you to better understand what activities there are in this gallery?
   ☐ Yes ☐ No ☐ Unsure
   a. *(If Yes)* How did they help?

Museum of Science. ICT

Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____

Condition: ☐ Direct ☐ Blended

8. What questions do you have about Ada and Grace *or* what else would you like to know about them?

9. Can you think of other settings where virtual humans might be useful?

10. What are your feelings about virtual humans, in general?

11. Next I have a few statements I'd like you to rate. When rating these, think about your experience with the exhibit as a whole [both the Twins and the area that explains about them *(if they visited Science Behind)*]. *(Show visual, circle their choice)*

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Being able to speak with the Twins was …	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting with the Twins would be…	Boring	Just okay	Pretty good	Exciting

12. On the scale of 1 to 4, how easy was it to figure out what to do here? *(Show visual, circle their choice)*

1) It was so easy I didn't have to think about it.	2) It was easy to figure out.	3) It was a little difficult to figure out.	4) It was so difficult I couldn't figure it out.

13. *(Adults only)* Now I have another list of statements I'd like you to rate. I will ask you to rate each statement twice: once thinking back to before you used the exhibit and again thinking about now that you have used it. Please rate each statement on a scale of 1-7, with 1 being "Strongly Disagree" and 7 being "Strongly Agree."

Before								After						
Disagree						Agree	Statement	Disagree						Agree
1	2	3	4	5	6	7	I understand what a virtual human is.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	I enjoy being able to speak to a computer as a way to interact with it.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	Having a computer with a personality is a good thing.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	In the future, there will be new and exciting innovations with smarter computers.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	In the future, interacting with computers will be easier.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	Women have made important contributions in the field of computer science.	1	2	3	4	5	6	7

Twins Interview Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____
ICT-MOS Summative Evaluation Condition: ☐ Direct ☐ Blended

Now I have just a couple of questions about you:

Thinking about you, how much would you say you know about… *(Show visual, circle their choice)*

Technology.	Nothing	A little	Quite a bit	A lot!
Computers.	Nothing	A little	Quite a bit	A lot!
Again **thinking about you**, how would you describe your interest is in these topics… *(Show visual, circle their choice)*				
Learning about technology is…	Boring	Just okay	Pretty good	Exciting
Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

(Do not ask, just mark one)

___ Female
___ Male

What is your zip code, if you live in the US?
___ ___ ___ ___ ___ (ex. 22602)

If in another country, please tell us what country: _____

What year were you born? _____

How many people are in your group today? _____

Are there any children in the group?
___ Yes
___ No

Do you work in or having training in one of the following fields: science, computers, or technology?
___ Yes
___ No

Follow-up questionnaire (Check the appropriate box)

☐ *(For those 16 and over):* We are interested in contacting visitors who have used this exhibit a few weeks from now. We would email you a short survey to complete. In exchange for completing the email survey, you will receive a $10 gift certificate to Amazon.com.

☐ *(For those under 16, address their parent):* We are interested in contacting visitors who have used this exhibit a few weeks from now. We would email **you** a short survey and ask you to complete it **with your child**. In exchange for completing the email survey, you will receive a $10 gift certificate to Amazon.com.

Are you interested in participating in the email survey?
☐ Yes ☐ No

(If Yes) May I have your email address?

_____

Museum of Science. ICT

Twins Follow-up Questionnaire
ICT-MOS Summative Evaluation

The contents of this questionnaire will be reproduced on Qualtrics. As such the appearance of the questions may vary, but the question order and content will be the same. Upon following the link in the either the invitation or reminder e-mail, participants will first see a page that describes the study purpose, length, and details of their participation. This information will appear as the first page before the participant begins the questionnaire. Clicking the "Start Questionnaire" button at the bottom of the page will both indicate their consent to participate and take them to the body of the questionnaire:

Questionnaire Version 1: Adult who experienced the Twins (either condition)

[Page 1]

You have been asked to take part in this questionnaire evaluation of the Virtual Humans Twins, Ada and Grace, at the Museum of Science Boston because you interacted with them during a recent visit to the museum and volunteered to be contacted. This study is designed to gather information on the impact of interacting with the Twins after visitors leave the museum. The museum is required to questionnaire visitors as part of the National Science Foundation grant that funded this project.

The questionnaire should only take about 10 minutes to complete and does not pose any risk and/or discomfort. **Your answers will be confidential.** No identifying information will be included in any reports resulting from this study. Your e-mail address and name will not be associated with your responses, and your IP address will not be collected during this questionnaire. Taking part in this study is completely voluntary and you are free to quit the questionnaire at any time.

If you have questions about this study or would like a copy of this consent page, please contact the director of the evaluation study:

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
(410) 956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

By clicking on the "Start Questionnaire" button you confirm that you have read the above information and are willing to take the questionnaire.

[Page 2]

[Questions 1 and 2- Overall satisfaction with their museum visit]

Based on your last visit, how would you rate the quality of the Museum of Science as an **educational experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all										Extremely

Educational Educational

Based on your last visit, how would you rate the quality of the Museum of Science as an **entertainment experience**?

0 1 2 3 4 5 6 7 8 9 10

Not at all Entertaining Extremely

[Question 3-remember the Twins?]

When you were at the museum, in the computer exhibit area, you interacted with Virtual Humans Twins, Ada and Grace. They appeared as life-size characters on a screen. Do you remember interacting with them?

- □ Yes [skip to Q 4]
- □ No [skip to Q 13]
- □ I don't know [Skip to Q 13]

[Page 3]

[Question 4-rating of interactions with the Twins]

For each of the following please select one word that describes your opinion. When rating these, think about your experience with both Ada and Grace and the area that explains about them.

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Being able to speak with the Twins was …	Boring	Just okay	Pretty good	Exciting
Learning more about how the Twins work would be…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting ith the Twins would be…	Boring	Just okay	Pretty good	Exciting

[Question 5-remember of their interaction with the Twins?]

Thinking back to you interaction with Ada and Grace and the area that explains about them, what do you remember most? [open-ended unlimited response]

[Question 6-technology that makes the Twins work]

Please list a few technologies you think are needed to make Ada and Grace work. [open-ended unlimited response]

[Page 4]

[Question 7-STEM learning]

What did you learn about technology or computers from interacting with the Twins? [open-ended unlimited response]

[Question 8- remember interaction with a staff member?]

When you were interacting with Ada and Grace, did you receive help from museum staff member (i.e. someone in a red coat or apron)?

- □ Yes [skip to Q 9]
- □ No [skip to Q 10]
- □ I don't know [Skip to Q 10]

[Page 5]
[Question 9- helpful about interaction with a staff member]
How did talking with a museum staff member help you to better understand Ada and Grace and how they work? [open-ended unlimited response]

[Page 6]
[Question 10-after their visit]
After you left the museum, did you talk to anyone about Ada and Grace or the exhibit explaining about them?

- ☐ Yes [skip to Q 11]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 7]
[Question 11-after visit who did they talk to?]
Who did you talk to about Ada and Grace after your visit? (check all that apply)

- ☐ A person who went to the museum with me that day
- ☐ A family member
- ☐ A friend or co-worker
- ☐ Someone else

[Question 12- subject of the conversation?]
What about Ada and Grace did you and the person talk about? [open-ended unlimited response]

[Page 8]
[Question 13-impact ratings]
Please rate each of the following statements. The rating scale is 1 = "Strongly Disagree" and 7 = "Strongly Agree."

Statement	Strongly Disagree						Strongly Agree
I understand what a virtual human is.	1	2	3	4	5	6	7
I enjoy being able to speak to a computer as a way to interact with it.	1	2	3	4	5	6	7
Having a computer with a personality is a good thing.	1	2	3	4	5	6	7
In the future, there will be new and exciting innovations with smarter computers.	1	2	3	4	5	6	7
In the future, interacting with computers will be easier.	1	2	3	4	5	6	7
Women have made important contributions in the field of computer science.	1	2	3	4	5	6	7

[Page 9]
[Question 14-personal knowledge/attitudes ratings]
You are almost done with the questionnaire. These are the last questions!

For each of the following, select one word that describes how much you know about the topic.

Technology.	Nothing	A little	Quite a bit	A lot!

Computers.	Nothing	A little	Quite a bit	A lot!

For each of the following please select one word that describes interest in the topic.

Learning about technology is…	Boring	Just okay	Pretty good	Exciting
Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

[Page 10]

Thank you for your help and feedback. This information will help the Museum of Science know whether they met their goals for the project.

As a thank you for completing the questionnaire, we'd like to offer you a $10 gift certificate to Amazon.com. Please include your email address in the space below if you would like receive the gift certificate. The email address will only be used for the purpose of emailing the gift certificate. It will not be used for any other purpose, shared or otherwise provided to any third parties.

Provide your email address here to receive the $10 gift certificate to Amazon.com:

Questionnaire Version 2: Child who experienced the Twins (either condition)

[Page 1]

You and your child have been asked to take part in this questionnaire evaluation of the Virtual Humans Twins, Ada and Grace, at the Museum of Science Boston because your child interacted with them during a recent visit to the museum and volunteered to be contacted. This study is designed to gather information on the impact of interacting with the Twins after visitors leave the museum. The museum is required to questionnaire visitors as part of the National Science Foundation grant that funded this project. **Please complete the questionnaire with your child.**

The questionnaire should only take about 10 minutes to complete and does not pose any risk and/or discomfort. **Your answers will be confidential.** No identifying information will be included in any reports resulting from this study. Your e-mail address and name will not be associated with your responses, and your IP address will not be collected during this questionnaire. Taking part in this study is completely voluntary and you are free to quit the questionnaire at any time.

If you have questions about this study or would like a copy of this consent page, please contact the director of the evaluation study:

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
(410) 956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

By clicking on the "Start Questionnaire" button you confirm that you have read the above information and are willing to take the questionnaire.

[Page 2]

Throughout this questionnaire "you" refers to the child who participated in the questionnaire at the museum. Adults, please help your child read and answer the questions—but keep in mind we really want the child's opinion!

[Questions 1 and 2- Overall satisfaction with their museum visit]

Based on your last visit, how would you rate the quality of the Museum of Science as an **educational experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Educational										Extremely Educational

Based on your last visit, how would you rate the quality of the Museum of Science as an **entertainment experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Entertaining										Extremely

[Question 3-remember the Twins?]

When you were at the museum, in the computer exhibit area, you interacted with Virtual Humans Twins, Ada and Grace. They appeared as life-size characters on a screen. Do you remember interacting with them?

- ☐ Yes [skip to Q 4]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 3]

[Question 4-rating of interactions with the Twins]

For each of the following please select one word that describes your opinion. When rating these, think about your experience with both Ada and Grace and the area that explains about them.

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Being able to speak with the Twins was …	Boring	Just okay	Pretty good	Exciting
Learning more about how the Twins work would be…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting ith the Twins would be…	Boring	Just okay	Pretty good	Exciting

[Question 5-remember of their interaction with the Twins?]

Thinking back to you interaction with Ada and Grace and the area that explains about them, what do you remember most? [open-ended unlimited response]

[Question 6-technology that makes the Twins work]

Please list a few technologies you think are needed to make Ada and Grace work. [open-ended unlimited response]

[Page 4]
[Question 7-STEM learning]
What did you learn about technology or computers from interacting with the Twins? [open-ended unlimited response]

[Question 8- remember interaction with a staff member?]
When you were interacting with Ada and Grace, did you receive help from museum staff member (i.e. someone in a red coat or apron)?

- ☐ Yes [skip to Q 9]
- ☐ No [skip to Q 10]
- ☐ I don't know [Skip to Q 10]

[Page 5]
[Question 9- helpful about interaction with a staff member]
How did talking with a museum staff member help you to better understand Ada and Grace and how they work? [open-ended unlimited response]

[Page 6]
[Question 10-after their visit]
After you left the museum, did you talk to anyone about Ada and Grace or the exhibit explaining about them?

- ☐ Yes [skip to Q 11]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 7]
[Question 11-after visit who did they talk to?]
Who did you talk to about Ada and Grace after your visit? (check all that apply)

- ☐ A person who went to the museum with me that day
- ☐ A family member
- ☐ A friend or co-worker
- ☐ Someone else

[Question 12- subject of the conversation?]
What about Ada and Grace did you and the person talk about? [open-ended unlimited response]

[Page 8]
[Question 13-personal knowledge/attitudes ratings]
You are almost done with the questionnaire. These are the last questions!

For each of the following, select one word that describes how much you know about the topic.

Technology.	Nothing	A little	Quite a bit	A lot!
Computers.	Nothing	A little	Quite a bit	A lot!

For each of the following please select one word that describes interest in the topic.

Learning about technology is…	Boring	Just okay	Pretty good	Exciting

Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

[Page 9]

Thank you for your help and feedback. This information will help the Museum of Science know whether they met their goals for the project.

As a thank you for completing the questionnaire, we'd like to offer you a $10 gift certificate to Amazon.com. Please include your email address in the space below if you would like receive the gift certificate. The email address will only be used for the purpose of emailing the gift certificate. It will not be used for any other purpose, shared or otherwise provided to any third parties.

Provide your email address here to receive the $10 gift certificate to Amazon.com:

Appendix 3 **Instruments: Coach Mike and Robot Park**

Coach Observation
ICT-MOS Summative Evaluation

Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____
Coach Mike Engaged *(circle one)*: Yes No

Time spent: _____ (min:sec)

Observed Behaviors

- ☐ Looks at the Screen *(When Mike is OFF)*
- ☐ Physically moves the robot
- ☐ Uses Block Tester
- ☐ Other: _____

Programs with Run pressed *(enter # of blocks)*

Attempted—Challenges—Completed

☐	Lights up sign	☐
☐	Turn around	☐
☐	Make a square	☐

Interacts with Coach Mike

- ☐ Mike button, # of times: _____
- ☐ Looks when talking
- ☐ Mentions to another visitor/staff member
- ☐ Tries to follow his direction/suggestion
- ☐ Other: _____

Usability Issues

- ☐ Camera
- ☐ Mat
- ☐ Frozen computer/crash
- ☐ Other:_____

Social Interactions

Target to Visitor

- ☐ Conceptual
- ☐ Operational
- ☐ Goal setting
- ☐ Positive Affect
- ☐ Negative Affect
- ☐ Other: _____

Target to Staff

- ☐ Conceptual
- ☐ Operational
- ☐ Goal setting
- ☐ Positive Affect
- ☐ Negative Affect
- ☐ Other: _____

Significant Interactions *(record here)*

Coach Observation
ICT-MOS Summative Evaluation

Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____
Coach Mike Engaged *(circle one)*: Yes No

Observed Demographics *(complete only in the absence of interview demographics)*

Target's gender:

- ☐ Female
- ☐ Male

Target's age range

- ☐ Under 18
- ☐ Over 18

Children in the Target's Group

- ☐ Yes, saw kids
- ☐ No kids, only saw adults

Coach Interview Date: _____ Time: _____ Data Collector: _____ Interview ID #:_____
ICT-MOS Summative Evaluation Coach Mike Engaged *(circle one)*: Yes No

Hi, the Museum is trying to get feedback about the exhibit you just used, and I was wondering if you would be willing to answer a few questions? It will take 5-10 minutes. FOR PARENTS OF CHILDREN: *Hi, the Museum is trying to get feedback about the exhibit your child just used, and I was wondering if I could interview your daughter/son? It will take 5-10 minutes.* IF YES: *Before we start, I want to let you both know that you don't have to answer the questions if you don't want to and that you can end the interview at any time. There are also no right or wrong answers to my questions. Any thoughts you share will help us improve our exhibits.*

1. Can you tell me what you did at the exhibit?

   a. Did anything you just did help you understand how computers work? (Probe: How did that help you understand more about computers? Can you tell me more about that?)

2. What would you say was the main idea of the exhibit? (Probe: What was the museum hoping visitors would learn?)

3. Next I have a few statements I'd like you to rate. When rating these, think about your experience with the exhibit as a whole: using the blocks, working with the robot, and interacting with Coach Mike *(if using the Coach Mike condition)*. *(Show visual, circle their choice)*

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Programming the robot to move was…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting with the exhibit would be…	Boring	Just okay	Pretty good	Exciting

4. On the scale of 1 to 4, how easy was it to figure out what to do here? *(Show visual, circle their choice)*

1) It was so easy I didn't have to think about it.	2) It was easy to figure out.	3) It was a little difficult to figure out.	4) It was so difficult I couldn't figure it out.

5. *(Adults only)* Now I have another list of statements I'd like you to rate. I will ask you to rate each statement twice: once thinking back to before you used the exhibit and again thinking about now that you have used it. Please rate each statement on a scale of 1-7, with 1 being "Strongly Disagree" and 7 being "Strongly Agree."

Before								After						
Disagree						Agree	Statement	Disagree						Agree
1	2	3	4	5	6	7	I can explain to my friends what it means to write a computer program.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	I know what it means to "debug" a program.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	In the future, there will be new and exciting innovations with smarter computers.	1	2	3	4	5	6	7
1	2	3	4	5	6	7	In the future, interacting with computers will be easier.	1	2	3	4	5	6	7

ICT-MOS Summative Evaluation

The contents of this questionnaire will be reproduced on Qualtrics. As such the appearance of the questions may vary, but the question order and content will be the same. Upon following the link in the either the invitation or reminder e-mail, participants will first see a page that describes the study purpose, length, and details of their participation. This information will appear as the first page before the participant begins the questionnaire. Clicking the "Start Questionnaire" button at the bottom of the page will both indicate their consent to participate and take them to the body of the questionnaire:

Questionnaire Version 1: Adult who experienced the Coach Mike condition

[Page 1]
You have been asked to take part in this questionnaire evaluation of the Robot Park exhibit at the Museum of Science Boston because you interacted with the exhibit during a recent visit to the museum and volunteered to be contacted. This study is designed to gather information on the impact of interacting with the Robot Park exhibit after visitors leave the museum. The museum is required to questionnaire visitors as part of the National Science Foundation grant that funded this project.

The questionnaire should only take about 5 minutes to complete and does not pose any risk and/or discomfort. **Your answers will be confidential.** No identifying information will be included in any reports resulting from this study. Your e-mail address and name will not be associated with your responses, and your IP address will not be collected during this questionnaire. Taking part in this study is completely voluntary and you are free to quit the questionnaire at any time.

If you have questions about this study or would like a copy of this consent page, please contact the director of the evaluation study:

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
(410) 956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

By clicking on the "Start Questionnaire" button you confirm that you have read the above information and are willing to take the questionnaire.

[Page 2]
[Questions 1 and 2- Overall satisfaction with their museum visit]
Based on your last visit, how would you rate the quality of the Museum of Science as an **educational experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Educational										Extremely Educational

Based on your last visit, how would you rate the quality of the Museum of Science as an **entertainment experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Entertaining										Extremely Entertaining

[Question 3-remember Robot Park?]
When you were at the museum, in the computer exhibit area, you interacted with the Robot Park exhibit. It was an area where you hooked together blocks shaped like puzzle pieces to make a robot move. Do you remember using the Robot Park exhibit?

- ☐ Yes [skip to Q 4]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 3]
[Question 4-rating of interactions with Robot Park]
For each of the following please select one word that describes your opinion. When rating these, think about your experience with using the blocks, working with the robot, and the Virtual Coach Mike who appeared on the TV screen.

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Programming the robot to move was…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting with the exhibit would be…	Boring	Just okay	Pretty good	Exciting

[Question 5-remember of their interaction at Robot Park?]
Thinking back to you interaction Robot Park, what do you remember most? [open-ended unlimited response]

[Question 6-STEM learning]
What did you learn about technology or computers from using interacting at Robot Park? [open-ended unlimited response]

[Page 4]
[Question 7-ability to describe programming]
The goal of Robot Park is to help you learning about programming. Programming a way you can give directions to a computer so it will do something—in this case to move a robot.

Below are a list of statements that may or may not describe programming. Using what you know about programming, please move each statement into the correct box.

Statements	Boxes
A series of small steps	Yes, this describes programming
Combining small steps into something bigger	No, does not describe programming
There is usually only one ways to solve a problem	
You might not be right the first time	
You might have to "debug" your program	
If you mess up, you can't fix it	

[Question 8- remember Coach Mike?]
When you were interacting with Robot Park there was a Virtual Coach on the TV screen at the exhibit. We call him Coach Mike. Do you remember receiving help from Coach Mike?

- ☐ Yes [skip to Q 9]
- ☐ No [skip to Q 10]
- ☐ I don't know [Skip to Q 10]

[Page 5]
[Question 9- helpful about interaction with Coach Mike]
How did Coach Mike help you to better understand how to program the robot at Robot Park? [open-ended unlimited response]

[Page 6]
[Question 10-after their visit]
After you left the museum, did you talk to anyone about the Robot Park exhibit?

- ☐ Yes [skip to Q 11]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 7]
[Question 11-after visit who did they talk to?]
Who did you talk to about Robot Park after your visit? (check all that apply)

- ☐ A person who went to the museum with me that day
- ☐ A family member
- ☐ A friend or co-worker
- ☐ Someone else

[Question 12- subject of the conversation?]
What about Robot Park did you and the person talk about? [open-ended unlimited response]

[Page 8]
[Question 13-impact ratings]
Please rate each of the following statements. The rating scale is 1 = "Strongly Disagree" and 7 = "Strongly Agree."

Statement	Strongly Disagree						Strongly Agree
I can explain to my friends what it means to write a computer program.	1	2	3	4	5	6	7
I know what it means to "debug" a program.	1	2	3	4	5	6	7
In the future, there will be new and exciting innovations with smarter computers.	1	2	3	4	5	6	7
In the future, interacting with computers will be easier.	1	2	3	4	5	6	7

[Page 9]
[Question 14-personal knowledge/attitudes ratings]
You are almost done with the questionnaire. These are the last questions!

For each of the following, select one word that describes how much you know about the topic.

Technology.	Nothing	A little	Quite a bit	A lot!
Computers.	Nothing	A little	Quite a bit	A lot!

For each of the following please select one word that describes interest in the topic.

Learning about technology is…	Boring	Just okay	Pretty good	Exciting
Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

[Page 10]

Thank you for your help and feedback. This information will help the Museum of Science know whether they met their goals for the project.

As a thank you for completing the questionnaire, we'd like to offer you a $10 gift certificate to Amazon.com. Please include your email address in the space below if you would like receive the gift certificate. The email address will only be used for the purpose of emailing the gift certificate. It will not be used for any other purpose, shared or otherwise provided to any third parties.

Provide your email address here to receive the $10 gift certificate to Amazon.com:

Questionnaire Version 2: Child who experienced the Coach Mike condition

[Page 1]
You and your child have been asked to take part in this questionnaire evaluation of the Robot Park exhibit at the Museum of Science Boston because your child interacted with the exhibit during a recent visit to the museum and volunteered to be contacted. This study is designed to gather information on the impact of interacting with the Robot Park exhibit after visitors leave the museum. The museum is required to questionnaire visitors as part of the National Science Foundation grant that funded this project. **Please complete the questionnaire with your child.**

The questionnaire should only take about 5 minutes to complete and does not pose any risk and/or discomfort. **Your answers will be confidential.** No identifying information will be included in any reports resulting from this study. Your e-mail address and name will not be associated with your responses, and

your IP address will not be collected during this questionnaire. Taking part in this study is completely voluntary and you are free to quit the questionnaire at any time.

If you have questions about this study or would like a copy of this consent page, please contact the director of the evaluation study:

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
(410) 956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

By clicking on the "Start Questionnaire" button you confirm that you have read the above information and are willing to take the questionnaire.

[Page 2]
Through out this questionnaire "you" refers to the child who participated in the questionnaire at the museum. Adults, please help your child read and answer the questions—but keep in mind we really want the child's opinion!

[Questions 1 and 2- Overall satisfaction with their museum visit]
Based on your last visit, how would you rate the quality of the Museum of Science as an **educational experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Educational										Extremely Educational

Based on your last visit, how would you rate the quality of the Museum of Science as an **entertainment experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Entertaining										Extremely

[Question 3-remember Robot Park?]
When you were at the museum, in the computer exhibit area, you interacted with the Robot Park exhibit. It was an area where you hooked together blocks shaped like puzzle pieces to make a robot move. Do you remember using the Robot Park exhibit?

- ☐ Yes [skip to Q 4]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 3]
[Question 4-rating of interactions with Robot Park]
For each of the following please select one word that describes your opinion. When rating these, think about your experience with using the blocks, working with the robot, and the Virtual Coach Mike who appeared on the TV screen.

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Programming the robot to move was…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting with the exhibit would be…	Boring	Just okay	Pretty good	Exciting

[Question 5-remember of their interaction at Robot Park?]
Thinking back to you interaction Robot Park, what do you remember most? [open-ended unlimited response]

[Question 6-STEM learning]
What did you learn about technology or computers from using interacting at Robot Park? [open-ended unlimited response]

[Page 4]
[Question 7-ability to describe programming]
The goal of Robot Park is to help you learning about programming. Programming a way you can give directions to a computer so it will do something—in this case to move a robot.

Below are a list of statements that may or may not describe programming. Using what you know about programming, please move each statement into the correct box.

Statements	Boxes
A series of small steps	Yes, this describes programming
Combining small steps into something bigger	No, does not describe programming
There is usually only one ways to solve a problem	
You might not be right the first time	
You might have to “debug” your program	
If you mess up, you can’t fix it	

[Question 8- remember Coach Mike?]
When you were interacting with Robot Park there was a Virtual Coach on the TV screen at the exhibit. We call him Coach Mike. Do you remember receiving help from Coach Mike?

- ☐ Yes [skip to Q 9]
- ☐ No [skip to Q 10]
- ☐ I don’t know [Skip to Q 10]

[Page 5]
[Question 9- helpful about interaction with Coach Mike]
How did Coach Mike help you to better understand how to program the robot at Robot Park? [open-ended unlimited response]

[Page 6]
[Question 10-after their visit]
After you left the museum, did you talk to anyone about the Robot Park exhibit?

- ☐ Yes [skip to Q 11]
- ☐ No [skip to Q 13]
- ☐ I don't know [Skip to Q 13]

[Page 7]
[Question 11-after visit who did they talk to?]
Who did you talk to about Robot Park after your visit? (check all that apply)

- ☐ A person who went to the museum with me that day
- ☐ A family member
- ☐ A friend or co-worker
- ☐ Someone else

[Question 12- subject of the conversation?]
What about Robot Park did you and the person talk about? [open-ended unlimited response]

[Page 8]
[Question 13-personal knowledge/attitudes ratings]
You are almost done with the questionnaire. These are the last questions!

For each of the following, select one word that describes how much you know about the topic.

Technology.	Nothing	A little	Quite a bit	A lot!
Computers.	Nothing	A little	Quite a bit	A lot!

For each of the following please select one word that describes interest in the topic.

Learning about technology is…	Boring	Just okay	Pretty good	Exciting
Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

[Page 9]

Thank you for your help and feedback. This information will help the Museum of Science know whether they met their goals for the project.

As a thank you for completing the questionnaire, we'd like to offer you a $10 gift certificate to Amazon.com. Please include your email address in the space below if you would like receive the gift certificate. The email address will only be used for the purpose of emailing the gift certificate. It will not be used for any other purpose, shared or otherwise provided to any third parties.

Provide your email address here to receive the $10 gift certificate to Amazon.com:

Questionnaire Version 3: Adult who did not experienced the Coach Mike condition

[Page 1]
You have been asked to take part in this questionnaire evaluation of the Robot Park exhibit at the Museum of Science Boston because you interacted with the exhibit during a recent visit to the museum and volunteered to be contacted. This study is designed to gather information on the impact of interacting with the Robot Park exhibit after visitors leave the museum. The museum is required to questionnaire visitors as part of the National Science Foundation grant that funded this project.

The questionnaire should only take about 5 minutes to complete and does not pose any risk and/or discomfort. **Your answers will be confidential.** No identifying information will be included in any reports resulting from this study. Your e-mail address and name will not be associated with your responses, and your IP address will not be collected during this questionnaire. Taking part in this study is completely voluntary and you are free to quit the questionnaire at any time.

If you have questions about this study or would like a copy of this consent page, please contact the director of the evaluation study:

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
(410) 956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

By clicking on the "Start Questionnaire" button you confirm that you have read the above information and are willing to take the questionnaire.

[Page 2]
[Questions 1 and 2- Overall satisfaction with their museum visit]
Based on your last visit, how would you rate the quality of the Museum of Science as an **educational experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Educational										Extremely Educational

Based on your last visit, how would you rate the quality of the Museum of Science as an **entertainment experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Entertaining										Extremely

[Question 3-remember Robot Park?]
When you were at the museum, in the computer exhibit area, you interacted with the Robot Park exhibit. It was an area where you hooked together blocks shaped like puzzle pieces to make a robot move. Do you remember using the Robot Park exhibit?

- ☐ Yes [skip to Q 4]
- ☐ No [skip to Q 11]
- ☐ I don't know [Skip to Q 11]

[Page 3]
[Question 4-rating of interactions with Robot Park]
For each of the following please select one word that describes your opinion. When rating these, think about your experience with using the blocks and working with the robot.

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Programming the robot to move was…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting with the exhibit would be…	Boring	Just okay	Pretty good	Exciting

[Question 5-remember of their interaction at Robot Park?]
Thinking back to you interaction Robot Park, what do you remember most? [open-ended unlimited response]

[Question 6-STEM learning]
What did you learn about technology or computers from using interacting at Robot Park? [open-ended unlimited response]

[Page 4]
[Question 7-ability to describe programming]
The goal of Robot Park is to help you learning about programming. Programming a way you can give directions to a computer so it will do something—in this case to move a robot.

Below are a list of statements that may or may not describe programming. Using what you know about programming, please move each statement into the correct box.

<u>Statements</u>
A series of small steps
Combining small steps into something bigger
There is usually only one ways to solve a problem
You might not be right the first time
You might have to "debug" your program
If you mess up, you can't fix it

<u>Boxes</u>
Yes, this describes programming
No, does not describe programming

[Page 5]
[Question 8-after their visit]
After you left the museum, did you talk to anyone about the Robot Park exhibit?

- ☐ Yes [skip to Q 9]
- ☐ No [skip to Q 11]
- ☐ I don't know [Skip to Q 11]

[Page 6]
[Question 9-after visit who did they talk to?]
Who did you talk to about Robot Park after your visit? (check all that apply)

- ☐ A person who went to the museum with me that day
- ☐ A family member
- ☐ A friend or co-worker
- ☐ Someone else

[Question 10- subject of the conversation?]
What about Robot Park did you and the person talk about? [open-ended unlimited response]

[Page 7]
[Question 11-impact ratings]
Please rate each of the following statements. The rating scale is 1 = "Strongly Disagree" and 7 = "Strongly Agree."

Statement	Strongly Disagree						Strongly Agree
I can explain to my friends what it means to write a computer program.	1	2	3	4	5	6	7
I know what it means to "debug" a program.	1	2	3	4	5	6	7
In the future, there will be new and exciting innovations with smarter computers.	1	2	3	4	5	6	7
In the future, interacting with computers will be easier.	1	2	3	4	5	6	7

[Page 8]
[Question 12-personal knowledge/attitudes ratings]
You are almost done with the questionnaire. These are the last questions!

For each of the following, select one word that describes how much you know about the topic.

Technology.	Nothing	A little	Quite a bit	A lot!
Computers.	Nothing	A little	Quite a bit	A lot!

For each of the following please select one word that describes interest in the topic.

Learning about technology is…	Boring	Just okay	Pretty good	Exciting
Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

[Page 9]

Thank you for your help and feedback. This information will help the Museum of Science know whether they met their goals for the project.

As a thank you for completing the questionnaire, we'd like to offer you a $10 gift certificate to Amazon.com. Please include your email address in the space below if you would like receive the gift certificate. The email address will only be used for the purpose of emailing the gift certificate. It will not be used for any other purpose, shared or otherwise provided to any third parties.

Provide your email address here to receive the $10 gift certificate to Amazon.com:

Questionnaire Version 4: Child who did not experienced the Coach Mike condition

[Page 1]
You and your child have been asked to take part in this questionnaire evaluation of the Robot Park exhibit at the Museum of Science Boston because your child interacted with the exhibit during a recent visit to the museum and volunteered to be contacted. This study is designed to gather information on the impact of interacting with the Robot Park exhibit after visitors leave the museum. The museum is required to questionnaire visitors as part of the National Science Foundation grant that funded this project. **Please complete the questionnaire with your child.**

The questionnaire should only take about 5 minutes to complete and does not pose any risk and/or discomfort. **Your answers will be confidential.** No identifying information will be included in any reports resulting from this study. Your e-mail address and name will not be associated with your responses, and your IP address will not be collected during this questionnaire. Taking part in this study is completely voluntary and you are free to quit the questionnaire at any time.

If you have questions about this study or would like a copy of this consent page, please contact the director of the evaluation study:

Susan Foutz
Senior Research Associate
Institute for Learning Innovation
foutz@ilinet.org
(410) 956-5144

The Institute for Learning Innovation has partnered with the Museum of Science to facilitate this questionnaire.

By clicking on the "Start Questionnaire" button you confirm that you have read the above information and are willing to take the questionnaire.

[Page 2]
Through out this questionnaire "you" refers to the child who participated in the questionnaire at the museum. Adults, please help your child read and answer the questions—but keep in mind we really want the child's opinion!

[Questions 1 and 2- Overall satisfaction with their museum visit]
Based on your last visit, how would you rate the quality of the Museum of Science as an **educational experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Educational										Extremely Educational

Based on your last visit, how would you rate the quality of the Museum of Science as an **entertainment experience**?

0	1	2	3	4	5	6	7	8	9	10
Not at all Entertaining										Extremely

[Question 3-remember Robot Park?]
When you were at the museum, in the computer exhibit area, you interacted with the Robot Park exhibit. It was an area where you hooked together blocks shaped like puzzle pieces to make a robot move. Do you remember using the Robot Park exhibit?

- ☐ Yes [skip to Q 4]
- ☐ No [skip to Q 11]
- ☐ I don't know [Skip to Q 11]

[Page 3]
[Question 4-rating of interactions with Robot Park]
For each of the following please select one word that describes your opinion. When rating these, think about your experience with using the blocks and working with the robot

Interacting with the exhibit was…	Boring	Just okay	Pretty good	Exciting
Programming the robot to move was…	Boring	Just okay	Pretty good	Exciting
Learning more about computers by interacting with the exhibit would be…	Boring	Just okay	Pretty good	Exciting

[Question 5-remember of their interaction at Robot Park?]
Thinking back to you interaction Robot Park, what do you remember most? [open-ended unlimited response]

[Question 6-STEM learning]
What did you learn about technology or computers from using interacting at Robot Park? [open-ended unlimited response]

[Page 4]
[Question 7-ability to describe programming]
The goal of Robot Park is to help you learning about programming. Programming a way you can give directions to a computer so it will do something—in this case to move a robot.

Below are a list of statements that may or may not describe programming. Using what you know about programming, please move each statement into the correct box.

Statements	Boxes
A series of small steps	Yes, this describes programming
Combining small steps into something bigger	No, does not describe programming
There is usually only one ways to solve a problem	
You might not be right the first time	
You might have to “debug” your program	
If you mess up, you can’t fix it	

[Page 5]
[Question 8-after their visit]
After you left the museum, did you talk to anyone about the Robot Park exhibit?

- ☐ Yes [skip to Q 9]
- ☐ No [skip to Q 11]
- ☐ I don’t know [Skip to Q 11]

[Page 6]
[Question 9-after visit who did they talk to?]
Who did you talk to about Robot Park after your visit? (check all that apply)

- ☐ A person who went to the museum with me that day
- ☐ A family member
- ☐ A friend or co-worker
- ☐ Someone else

[Question 10- subject of the conversation?]
What about Robot Park did you and the person talk about? [open-ended unlimited response]

[Page 7]
[Question 11-personal knowledge/attitudes ratings]
You are almost done with the questionnaire. These are the last questions!

For each of the following, select one word that describes how much you know about the topic.

Technology.	Nothing	A little	Quite a bit	A lot!
Computers.	Nothing	A little	Quite a bit	A lot!

For each of the following please select one word that describes interest in the topic.

Learning about technology is…	Boring	Just okay	Pretty good	Exciting
Figuring out how computers work is…	Boring	Just okay	Pretty good	Exciting

Thank you for your help and feedback. This information will help the Museum of Science know whether they met their goals for the project.

As a thank you for completing the questionnaire, we'd like to offer you a $10 gift certificate to Amazon.com. Please include your email address in the space below if you would like receive the gift certificate. The email address will only be used for the purpose of emailing the gift certificate. It will not be used for any other purpose, shared or otherwise provided to any third parties.

Provide your email address here to receive the $10 gift certificate to Amazon.com:

Additional Follow-up Online Questionnaire Findings

In the follow-up questionnaire, respondents were asked to rate the quality of the Museum of Science as an educational experience with a scale of 0=*"not at all educational"* and 4=*"extremely educational."* Ratings ranged from 4 to 10, with a median rating of 9 (Table 86). Overall, respondents rated the museum as highly educational. Twenty-one people, 36% of respondents, rated the educational experience a 10, indicating the museum was extremely educational. There is no statistical evidence to support a difference in medians between the treatment and control group (Mann-Whitney U=418.5, N=59, p=0.831). In other words, both groups found the Museum of Science to be an educational experience.

Table 93: Twins Follow-up Rating for Museum Educational Experience

Stay Time	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	59	8.59	9	1.464	4	10
Direct	32	8.63	9	1.476	4	10
Blended	27	8.56	9	1.476	4	10

Respondents were also asked to rate the quality of the Museum of Science as an entertainment experience based on their last visit. The average rating was 8.83 with a median of 9, indicating that respondents found the Museum of Science to be a very entertaining experience. When ratings were compared by interaction type (Blended or Direct), no significant differences were discovered (Mann-Whitney U=370.5, N=59, p=0.329).

Table 94: Twins Follow-up Rating for Museum Entertainment Experience

Stay Time	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	59	8.83	9	1.147	6	10
Direct	32	8.72	9	1.114	7	10
Blended	27	8.96	9	1.192	6	10

When asked if they remembered the Virtual Human Twins, all but one respondent remembered interacting with Ada and Grace (Table 88). The one respondent who marked "I don't know" was part of the Direct Interaction group.

Table 95: Responses to "*Do you remember interacting with the Twins*?" (n=59)

Remember	n	Percent
Yes	58	98%
No	0	0%
I don't know	1	2%

Thinking about their experience with both Ada and Grace, respondents in the follow-up questionnaire rated their opinion of the following statements:

- Interacting with the exhibit was…
- Being able to speak with the Twins was…
- Learning more about how the Twins work would be…
- Learning more about computers by interacting with the Twins would be…

Values were assigned to each rating: Boring =1; Just Okay=2; Pretty Good=3; Exciting=4. The median rating was 3 for each statement. Respondents felt that interacting with the exhibit, being able to speak with the Twins, learning about how the Twins work, and learning more about computers was "pretty good." Based on ratings from the follow-up questionnaire, visitors seemed to be really interested in the Twins, with the majority responding with "pretty good" or "exciting" for each statement. There were no significant differences when comparing ratings between the Direct and Blended Interaction conditions.

Table 96: Twins Interest Ratings from Follow-up Questionnaire (n=58)

	What is your opinion of:			
	Interacting with the exhibit	Being able to speak with the Twins	Learning more about how the Twins work	Learning more about computers
Mean Rating	3.12	3.22	3.14	3.09
Boring	2%	3%	3%	2%
Just okay	16%	9%	14%	22%
Pretty good	52%	50%	48%	41%
Exciting	31%	38%	35%	35%

Researchers also looked at how visitors' opinions of the Twins exhibit affected overall ratings for the Museum of Science. Several significant findings emerged.

- There is a weak positive correlation between opinions of interacting with the Twins exhibit and the overall educational experience at the Museum of Science (Spearman's correlation= 0.346, N=58, *p*=0.008). The same holds true for the entertainment rating (Spearman's

correlation= 0.298, N=58, *p*=0.023). A more positive rating for interacting with the exhibit will result in a higher overall rating for educational and entertainment experience.

- Visitors who enjoyed being able to speak with the Twins also felt the Museum of Science provided a more educational experience (Spearman's correlation= 0.339, N=58, *p*=0.009).
- A more positive rating for learning about computers directly correlates with a visitors overall rating for the Museum of Science as an educational experience (Spearman's correlation= 0.510, N=58, *p*=0.000). This also holds true for rating the entertainment experience of the Museum of Science (Spearman's correlation= 0.474, N=58, *p*=0.000).

As visitors were asked these same questions in the on-site interview, researchers compared ratings from the interview and follow-up questionnaire. Ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Follow-up questionnaire participants were asked the closed-ended question: "When you were interacting with Ada and Grace, did you receive help from museum staff member (i.e., someone in a red coat or apron)?" Almost three-quarters of the respondents (69%) indicated they received help from a museum staff member. About 24% indicated they did not receive help from a staff memebr, and 7% did not recall whether they had received help from a staff member at the exhibit. Blended participants were more likely to **receive help from a museum staff member** ($\chi 2$ = 18.412, N=58, *p*=0.000). Of the 27 participants in the blended category, 26 reported receiving help from a museum staff member, compared to 14 of the 31 Direct participants.

Follow-up questionnaire participants who agreed that a museum staff member helped them were asked the open-ended question: "How did talking with a museum staff member help you to better understand Ada and Grace and how they work?" The majority (45%) said the staff person explained how the twins worked, explained their design and technology, and potential future uses for the technology. Almost as many respondents (43%) reported that staff explained how to use logistical aspects of the exhibit. Participants under the age of 16 were more likely to receive help from a museum staff member **regarding how to use the exhibit** than adult participants ($\chi 2$ = 4.220, N=37, *p*=0.040). About 13% of respondents said that the staff members were helpful in general. Some examples of the comments received included:

If the staff member wasn't present, we probably would have bypassed Ada and Grace.
He was able to help me speak to them without confusing them!
It helped me understand the people behind the technology.

Appendix 5 Additional Analyses: Coach Mike and Robot Park

Additional Onsite Interview Findings

The following results are statistically significant but not practically significant. Researchers are unsure of the meaning for these findings.

Response to the question "Can you tell me what you did at the exhibit?" by Social Interactions

Made robot move	x	Overall visitor interaction	(χ^2= 6.051, N=223, p = 0.014)
Interact with Mike or staff	x	Overall visitor interaction	(χ^2= 5.809, N=223, p = 0.016)
Interact with Mike or staff	x	Operational visitor interaction	(χ^2= 8.056, N=223, p = 0.005)
Made robot move	x	Goal setting visitor interaction	(χ^2= 6.538, N=223, p = 0.011)
Put blocks together	x	Operational staff interaction	(χ^2= 5.181, N=223, p = 0.023)
"Program"	x	Conceptual staff interaction	(χ^2= 8.036, N=223, p = 0.005)

Additional Follow-up Online Questionnaire Findings

The follow-up questionnaire asked respondents to rate the quality of the Museum of Science as an educational experience, with a scale of 0=*"not at all educational"* and 4=*"extremely educational."* The median rating for educational experience was a 9 meaning that respondents perceived the museum as highly educational (Table 90). There is no statistical evidence to support a difference in medians between the treatment and control group (Mann-Whitney U=569, N=73, p=0.282).

Table 97: Robot Park Follow-up Rating for Museum Educational Experience

Stay Time	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	73	8.33	9	1.463	3	10
Treatment	34	8.53	9	1.376	5	10
Control	39	8.15	8	1.531	3	10

Respondents were also asked to rate the quality of the Museum of Science as an entertainment experience, with a scale of 0=*"not at all entertaining"* and 4=*"extremely entertaining."* Ratings ranged from 5 to 10, with a median rating of 9 (Table 91). Overall, respondents found the Museum of Science to be very entertaining. Twenty-two people, 30% of respondents, rated the entertainment experience a 10, indicating the museum was extremely entertaining. There was no significant difference between the treatment and control group's median rating (Mann-Whitney U=548.5, N=73, p=0.188). Those that attended Robot Park found the museum to be very entertaining regardless of whether or not Coach Mike was engaged.

Table 98: Robot Park Follow-up Rating for Museum Entertainment Experience

Stay Time	n	Mean	Median	St. Dev.	Min	Max
Overall Exhibit	73	8.60	9	1.222	5	10
Treatment	34	8.82	9	1.141	6	10
Control	39	8.41	9	1.272	5	10

As the follow-up questionnaire was conducted six weeks after a participant's visit to the museum, respondents were provided with a brief description of Robot Park and asked if they remember using the Robot Park exhibit. As shown in Table 92, almost everyone remembered using the Robot Park exhibit (97%).

Table 99: Responses to "*Do you remember using the Robot Park exhibit*?" (n=73)

Remember	n	Percent
Yes	71	97%
No	1	1%
I don't know	1	1%

Thinking about their experience using the blocks, working with the robot, and Virtual Coach Mike, respondents rated their experience in Robot Park. For each of the following statements respondents were asked to select one word (Boring =1; Just okay=2; Pretty good=3; Exciting=4) that describes their opinion:

- Interacting with the exhibit was…
- Programming the robot to move was…
- Learning more about computers by interacting with the exhibit would be…

The average response was "pretty good" for all three statements (Table 93). Over half of the respondents (53%) found programming the robot to move exciting.

Table 100: Robot Park Interest Ratings from Follow-up Questionnaire (n=70)

	What is your opinion of:		
	Interacting with the exhibit	Programming the robot to move	Learning more about computers
Mean Rating	3.24	3.49	3.2
Boring	0%	0%	1%
Just okay	9%	4%	16%
Pretty good	59%	43%	44%
Exciting	32%	53%	39%

When comparing the exhibition experience ratings from the follow-up questionnaire between the treatment and control group, no significant differences emerged. Researchers also compared respondents' ratings for the Museum of Science as both an educational and entertainment experience to their exhibition experience ratings. There is a weak positive correlation between respondents viewing the Museum of Science as an entertaining experience and interacting with Robot Park (Spearman's correlation= 0.312, N=70, *p*=0.009). The same holds true for programming the robot to move (Spearman's correlation= 0.303, N=70, *p*=0.011). Respondents who rated their experience interacting with Robot Park or programming the robot to move higher were more likely to find the Museum of Science as a highly entertaining museum.

Table 94 shows visitors' exhibition experience ratings, during their interview on-site and in the follow-up questionnaire. Visitors' perception of their experience at Robot Park did not significantly change six weeks after their visit (Wilcoxon Signed Rank Test). Thus, interest held over time.

Table 101: Comparing Robot Park Experience Ratings

	Interview			Follow-up questionnaire		
Scale	n	Mean	Median	n	Mean	Median
Interacting with the exhibit	75	3.20	3	70	3.24	3
Programming the robot to move	74	3.39	4	70	3.49	4
Learning more about computers	75	3.04	3	70	3.20	3

Visitors who interacted with Robot Park when Coach Mike was operating were asked additional questions in the follow-up questionnaire about Coach Mike. The majority of respondents indicated that they remembered Coach Mike (75%).

Table 102: Remember Coach Mike (n=32)

Correctly Identified	n	Percent
Yes	24	75%
No	7	22%
I don't know	1	3%

Respondents were asked what they remembered most about Robot Park. Almost half (45%) of respondents mentioned making the robot move or perform an action, 38% referred to putting blocks together, 19% remembered trying to figure out the exhibit, 16% expressed a positive affect, and 14% discussed completing a challenge or task. Only 9% of respondents mentioned "programming" a robot. Other responses generally mentioned robots, technology or computers.

Table 103: Memories of Robot Park (n=64)

Responses	n	Percent*
Made the robot move or "do something"	29	45%
Put blocks/puzzle pieces together	24	38%
Figuring it out	12	19%
Positive affect	10	16%
Attempted/accomplished task	9	14%
"Program" a robot	6	9%
"Play" or general mentions of the robot	3	5%
Other	8	13%

* Multiple responses allowed. Percents may add up to more than 100.

Made the robot move or "do something": The most commonly remembered aspect of the exhibit focused on controlling the robot. These respondents described making the robot move or make noises. Some examples of respondents' comments include:

Figuring out how to make the robot move.
Making it [the robot] growl a lot.
Getting the robot through the blocks.

Put blocks/puzzle pieces together: Over a third of respondents recalled using blocks or puzzle pieces to move the robot. Responses included "I remember putting the pieces together and making the robot move" and "trying to figure out why the robot couldn't read puzzle pieces past a certain point."

Responses did not vary between treatment and control groups (Chi-Square test). One significant finding emerged when comparing responses of adults and children. Adults were significantly more likely to remember the challenge of figuring out the exhibit (χ^2= 8.862, N=64, p=0.003). It is important to note that the follow-up questionnaire specifically reminded adults that "throughout the questionnaire 'you' refers to the child who participated in the questionnaire at the museum." Although adults were asked to record the child's opinion, we cannot be certain that responses accurately represent the child's thoughts.

Appendix 6 Impacts and Indicators

Finalized Impacts and Indicators (Updated 6.3.2011)

Impact 1: Awareness (Yes/Maybe=Team's assessment on whether the indicator applies to the exhibit)

Method	Impact (Shaded) and its Related Indicators	Twins and Science Behind	Coach Mike
	Children (ages 7 -14) and adults will increase their awareness about computer science and technology.	Yes	Yes
Retrospective Pre, Post, Delayed Post	Visitors will increase their awareness of what a virtual human is.	Yes	Maybe (will not be evaluated)
Post	Visitors will be able to describe Ada & Grace as a computer that acts like a human.	Yes	
Post	Visitors will recognize at least one of the following as characteristics of the Twins: a) they interact through speech, b) they are able to respond, c) their responses are dependent on what is asked of them, d) they have non-verbal behaviors.	Yes	
Retrospective Pre, Post	Visitors will increase their awareness of the role of women as role models in computer science.	Yes	
Post	Visitors will recognize Ada and Grace as relating to the objectives of Cahners ComputerPlace, with the following indicators of awareness: 1) they highlight the same subjects as Cahners (computers, communications, robots), 2) they are "guides" to the space, directing visitors to other exhibits in Cahners.	Yes	
Post	Visitors will recognize at least one of the following as characteristics of Coach Mike: a) his responses are tied to what the visitor is doing, b) he is trying to help them problem-solve.		Yes

Impact 2: Engagement and Interest (Yes/Maybe=Team's assessment on whether the indicator applies to the exhibit)

Method	**Impact (Shaded) and its Related Indicators**	**Twins and Science Behind**	**Coach Mike**
	Children (ages 7 -14) and adults will increase their engagement and interest in computer science and technology.	Yes	Yes
Post, Delayed Post	Visitors will indicate that they had a positive experience at the exhibit.	Yes	Yes
Retrospective Pre, Post, Delayed Post	Visitors will indicate their interest in learning more about computer science.	Yes	Yes
Delayed Post	Visitors will have a conversation after they leave the museum about the exhibit experience.	Yes	Yes
Post, Delayed Post	Visitors will indicate an interest to learn more about at least one of the following aspects: a) how the Twins work, b) other examples of virtual humans, c) other uses for virtual human technology.	Yes	
Observation	Visitors will have conversations with other visitors, members of their group, or MOS staff about how the Twins work.	Yes	
Observation	Visitors to Robot Park will engage more deeply when Coach Mike is turned on, with the following indicators of engagement: 1) stay time at the exhibit, 2) number of programs created, 3) length of programs created, 4) completion of a task/goal incorporated into the exhibit.		Yes

Impact 3: Knowledge (Yes/Maybe=Team's assessment on whether the indicator applies to the exhibit)

Method	Impact (Shaded) and its Related Indicators	Twins and Science Behind	Coach Mike
	Children (ages 7 -14) and adults will increase their knowledge about computer science and technology.	Yes	Yes
Post, Delayed Post	Visitors will be able to name at least one aspect of what makes up virtual humans, such as: a) communications technology, b) artificial intelligence, c) natural language, d) animation/graphic, or e) nonverbal behavior.	Yes	Maybe (will not be evaluated)
Post, Delayed Post	Visitors will be able to discuss at least one technology that is needed to build a virtual human.	Yes	Maybe (will not be evaluated)
Post, Delayed Post	Visitors will learn at least one new idea related to the STEM domain they talked to the Twins about.	Yes	
Post, Delayed Post	Visitors can describe that computers are programmed by breaking large programs/tasks into smaller steps.		Yes
Post, Delayed Post	Visitors to Robot Park will demonstrate an increase in their knowledge of programming concepts (i.e. the nature of programming) when Coach Mike is turned on, as indicated at least one of the following: a) programming as a series of steps, b) programs are rarely correct the first time, c) revisions are often necessary, and d) failure is not the end of the process.		Yes

Impact 4: Attitude (Yes/Maybe=Team's assessment on whether the indicator applies to the exhibit)

Method	Impact (Shaded) and its Related Indicators	Twins and Science Behind	Coach Mike
	Children (ages 7 -14) and adults will have a positive attitude about computer science and technology.	Yes	Yes
Retrospective Pre, Post, Delayed Post	Visitors will indicate they like learning about technology.	Yes	Yes
Post, Delayed Post	Visitors will have a positive attitude towards virtual humans in society, as indicated by having positive perceptions of: a) interacting with a computer, b) virtual humans having a presence and persona, c) using natural language with a computer, and d) future developments in the field of artificial intelligence.	Yes	Maybe
Post	Visitors will enjoy interacting with the Virtual Humans.	Yes	Yes
Retrospective Pre, Post, Delayed Post	Visitors will indicate they like programming or working with robots.		Yes
Post	Visitors to Robot Park will be less frustrated with the experience when Coach Mike is turned on.		Yes
Post	Visitors will describe Coach Mike as either: a) a positive influence (supportive, helpful, motivating) or b) recognize his suggestions as helpful or useful.		Yes