

***Linking Evidence to Explanation in Global Science:***  
**Evaluation Report**

March 14, 2011

Center for Research, Evaluation, & Assessment  
The Lawrence Hall of Science  
University of California, Berkeley

Maia Werner-Avidon  
Celia Castillo  
Lisa Newton  
Dawn Robles  
Scott Randol

## INTRODUCTION

*Linking Evidence to Explanation in Global Science* is a project of the Lawrence Hall of Science designed to develop curriculum-based content modules for spherical display systems (e.g. Science on a Sphere) that engage museum visitors in the process of observing and interpreting patterns of global climate data. As part of this project, the Lawrence Hall of Science developed two content modules on the topics of ocean currents and weather. Each of these content modules was designed to operate in three different modes: 1) autorun; 2) facilitated by instructor or docent; and 3) hands-on workshop. The goals of the project were to engage visitors in interpreting scientific data, recognizing patterns in data, and, ultimately, using that data as evidence to support them in formulating explanations related to earth science content.

The content modules developed by the Lawrence Hall of Science included:

- **Go with the Flow:** This content module focused on the movement of ocean currents and how objects dropped in the ocean (e.g. Nike tennis shoes or a penguin feather) would move based on those currents. Offered as a single presentation in the autorun and workshop modes, this module was divided into two presentations for the docent-led presentation (Go with the Flow—Shoes and Go with the Flow—Penguin).
- **Global Weather:** This content module focused on front formation and weather prediction.

Evaluation of the exhibit was conducted by the Center for Research, Evaluation, & Assessment (REA) at the Lawrence Hall of Science. The primary goal of the evaluation was to look specifically at whether or not participating visitors used scientific thought processes (e.g. using evidence to support their claims, identifying patterns, presenting alternative ideas) during the presentations.

Evaluation questions included:

- Does participation in the global weather and ocean currents presentations on the Science on a Sphere encourage visitors to use evidence and reasoning to back up claims? Are visitors able to apply the use of evidence and reasoning when presented with a new, but related situation?
- What other patterns of scientific thought are exhibited during the presentation (e.g. identification of patterns, presentation of alternative ideas)?
- Is there a difference in the use of scientific thought processes depending on the presentation mode (autorun, docent-led presentation, workshop)?
- Does the program work well for a variety of age ranges?

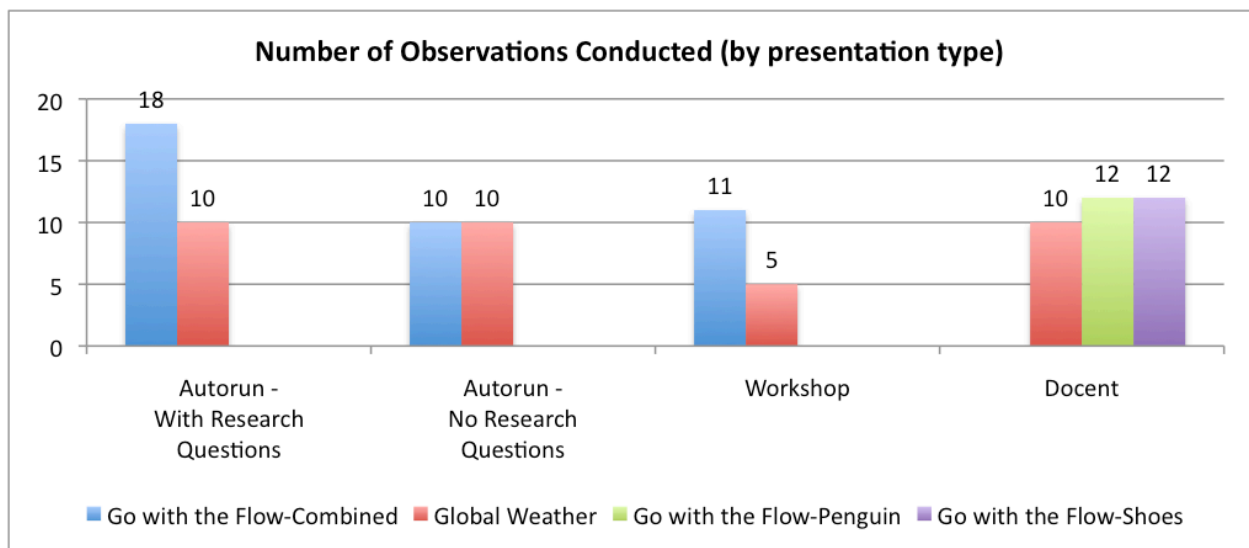
## METHODS

In an effort to answer these questions, evaluators from REA used two evaluation methods: visitor observations and written research questions.

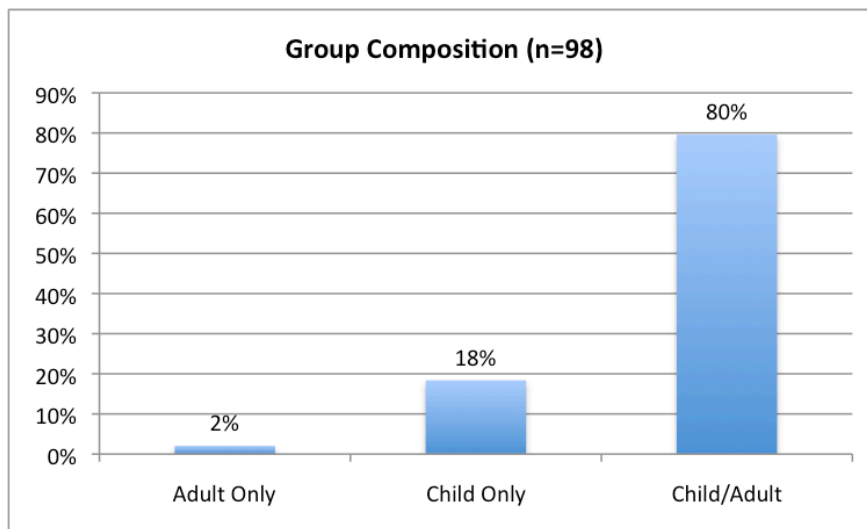
### *Visitor Observations:*

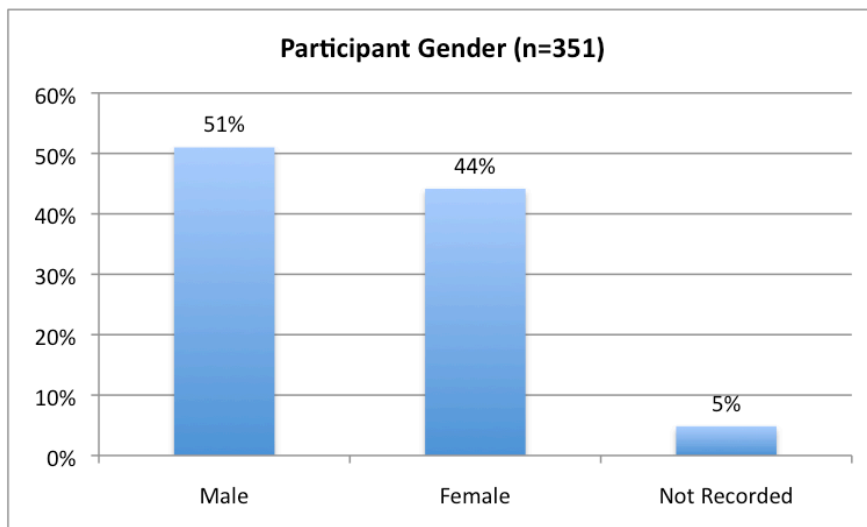
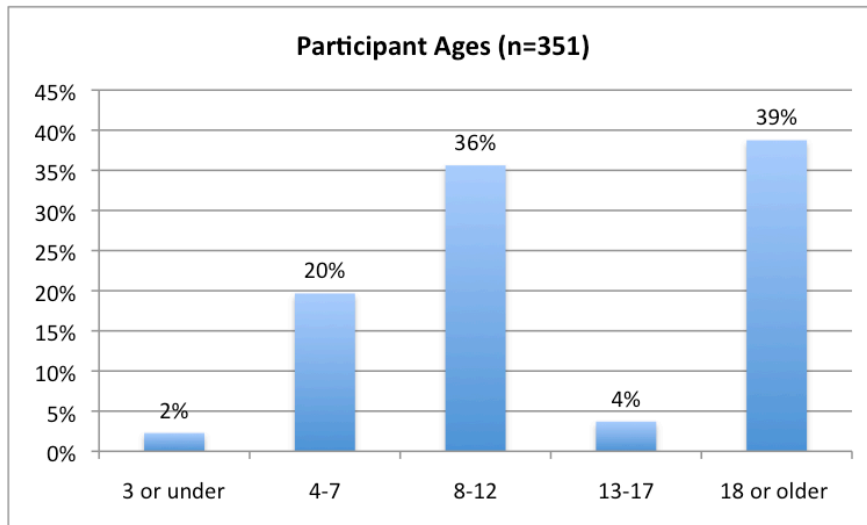
Staff from REA conducted observations of 98 groups watching the LEEGS presentations. Group size averaged 3.58 individuals per group, for a total of 351 individuals. Observations were used to examine the verbal communication taking place during the presentation (e.g. identification of

patterns, use of evidence and explanation verbally, presentation of alternative ideas, discussion within and across groups). Sixteen (16) observations were conducted of hands-on workshops (11 of these were for the Go with the Flow module and 5 for the Global Weather module). Thirty-four (34) observations were conducted for the docent-led presentations (12 for Go with the Flow—Shoes; 12 for Go with the Flow—Penguin; and 10 for Global Weather). Forty-eight (48) observations were conducted of the autorun version of the program. Twenty-eight (28) of these were autorun presentations in which the presentation was stopped to have visitors complete the written research questions (18 for the Go with the Flow module and 10 for the Global Weather module). Then, because researchers felt that the use of the written research questions might be impacting visitor behavior during the autorun presentations, an additional 20 autorun observations were conducted without the use of the written research questions (10 for Go with the Flow and 10 for Global Weather).



Demographic information about the observed visitors is detailed below:

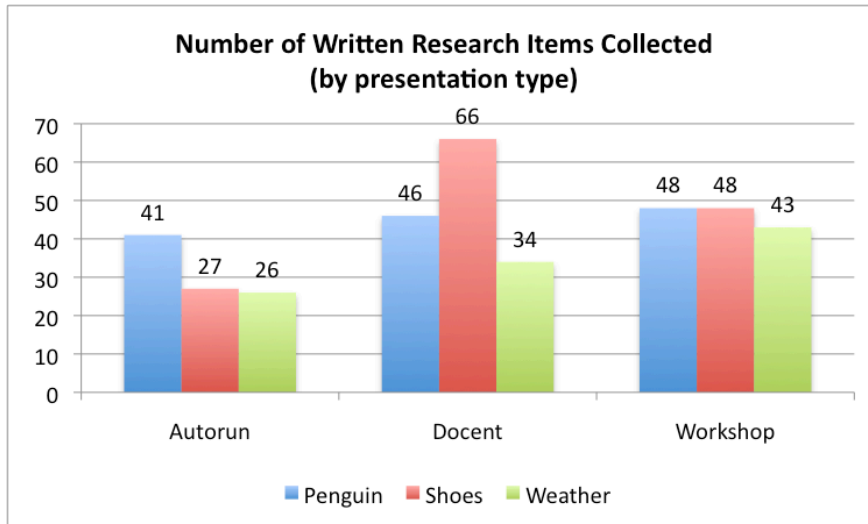




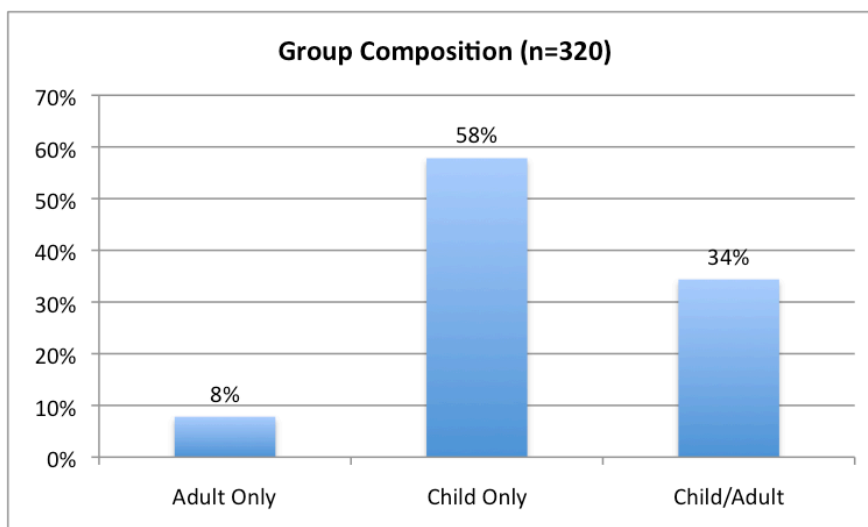
*Written Research Questions:*

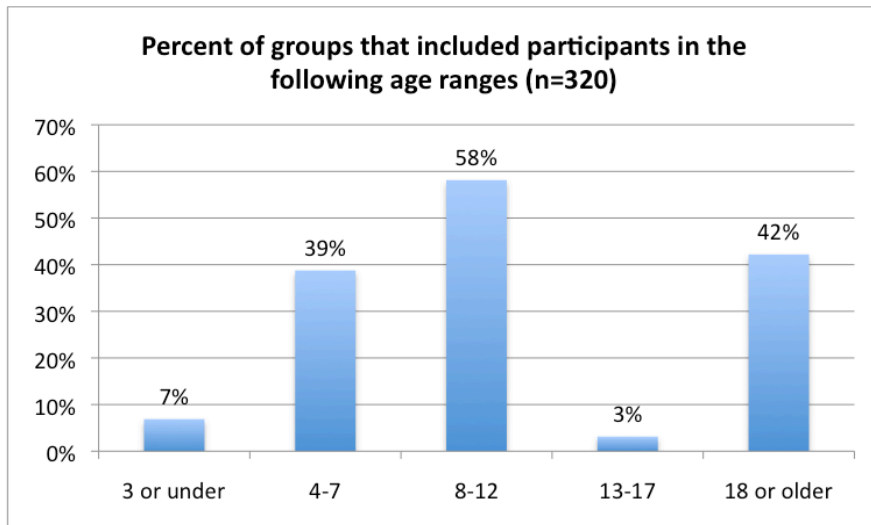
A major challenge to this evaluation study was that visitor observation alone would not necessarily be sufficient to understand visitors' use of evidence to support their claims because the amount of verbal interaction varies greatly depending on the format. Autorun, for example, is not necessarily conducive to verbal communication, but this does not necessarily mean that visitors are not using evidence to make explanations that are not verbally expressed. REA responded to this challenge by utilizing a written research instrument, in addition to the observations. With the written instrument, visitors were asked to write down a claim at two points in the presentation. First, visitors are asked to write a claim that is directly related to the evidence they have seen in the presentation. Then, visitors are asked to write down a claim that is not directly related to the presentation, but that would use similar evidence. Visitors' written items were analyzed to evaluate the extent to which visitors used evidence to support their claims. Data was also analyzed to examine differences in the presentation formats. Written research items were collected from a total of 379 visitor groups. For the autorun presentations, research items were collected from 94 groups (41 for Go with the Flow—Penguin, 27 for Go with the Flow—Shoes, and 26 for Global Weather). (Note that although the autorun presentation

included both the Shoes and Penguin presentations, visitors were only asked to write responses for one part of the presentation). For the docent presentations, research items were collected from 146 groups (46 for Penguin, 66 for Shoes, and 34 for Weather). For the workshops, research items were collected from 139 groups (48 for Penguin, 48 for Shoes, and 43 for Weather). (In the Go with the Flow workshops, participants were asked to write responses for both the Penguin and the Shoes research questions).



Participants were asked to answer the written research questions in small groups (family groups for autorialun or docent-led presentation and small groups of school children for the workshops). Group size averaged 2.56 individuals per group. Demographic information about the participants is detailed below (note: participants were not asked their gender for the research items):





Note that for the observations, we collected data on the number of participants in each age range. For the research questions, we asked participants to indicate the age ranges of people that worked on the research questions and the total number of individuals who worked on answering the questions, but did not ask them to indicate the number of people in each age range. As a result the data on age range is somewhat different for the research questions than for the observations.

## SUMMARY OF FINDINGS

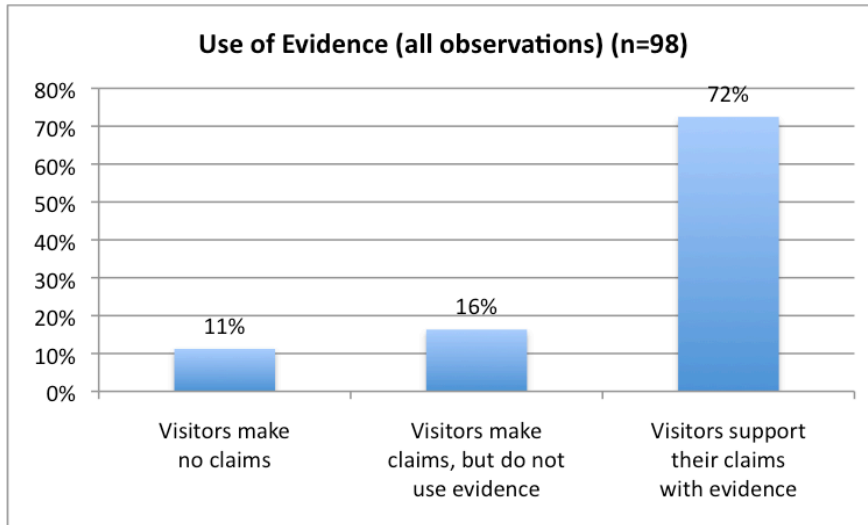
The following summarizes the key findings from the program evaluation. Each of these findings is outlined in greater detail below.

- Overall, the LEEGS programs were successful in getting visitors to use evidence to support their claims about what would happen on the sphere.
- The LEEGS programs were somewhat less successful at getting visitors to use other patterns of scientific thought (e.g. identification of patterns, presentation of alternative ideas), although they were extremely successful at getting visitors to engage in discussion.
- Workshop and docent-led presentations are more successful than autorun at engaging visitors in verbally expressing scientific thought processes. However, it appears that the use of some scientific thought processes can be increased in the autorun mode simply by stopping the autorun presentation and providing visitors with a chance to work together to express their ideas about what is happening in the presentation.
- Workshop and docent-led presentations appear to be more successful than autorun at getting visitors to use moderate or detailed evidence in supporting their explanations. However, this finding does not hold true across all three of the different presentations.
- Groups where all the children in the group are 8 or older are more successful than groups with younger children at using moderate or detailed evidence in supporting their explanations.

## DETAILED FINDINGS

*Finding 1: Overall, the LEEGS programs were successful in getting visitors to use evidence to support their claims about what would happen on the sphere.*

When all the observations are examined together (regardless of specific presentation or presentation type), it appears that the programs were very successful in engaging visitors in using evidence to support their claims. Seventy-two percent (72%) of observed visitors verbally used some kind of evidence in making a claim about what would happen in the presentation. For the Go with the Flow presentation, use of evidence typically included pointing at the sphere, discussing the direction of the currents, or using a laser pointer to show the path of the currents that the shoes or penguin feather would follow. For the Global Weather presentation, use of evidence typically included discussions of cloud patterns and the location of fronts and use of the time information to predict when the rain would hit California.



Similarly, the written research questions revealed high use of evidence to support claims. As noted above, participants were asked to respond to two written items. First, visitors are asked to write a claim that is directly related to the evidence they have seen in the presentation. Then, visitors are asked to write down a claim that is not directly related to the presentation, but that would use similar evidence. The specific questions for each presentation are outlined below:

*Global Weather*

- 1) Do you think it is going to rain during the concert? Why or why not?
- 2) Do you think it is going to rain in the next few days? Why or why not?

*Go with the Flow—Shoes*

- 1) Could the shoes get to Hawaii? Why or why not?
- 2) If the containers had fallen off near Manila, would the shoes have been able to get to Hawaii? Why or why not?

*Go with the Flow—Penguin*

- 1) Could a penguin feather from Antarctica get to California? Why or why not?
- 2) Could a puffin feather from Alaska get to California? Why or why not?

Responses were scored on a 5-point scale that included the following categories:

1 - Claim only (Yes/No or indicates what they think will happen without any explanation or evidence)

- “No, they would go to Oregon to San Francisco” (Shoes)
- “We think that would be possible” (Penguin)
- “I think it is going to rain at the concert” (Global Weather)

2 - Claim with unrelated evidence (offers explanation/evidence, but not related to what they see on the sphere)

- “No, they would get stuck in a ship’s engine and sink it” (Shoes)
- “Yes because it could fly to SF” (Penguin)
- “Yes. Because at the rate of the wind at the coast can tend to be strong.” (Global Weather)

3 - Minimal evidence (describes mechanism that would carry object or cause rain, but does not specify further)

- “Yes, because of the ocean currents” (Shoes)
- “Yes, if it followed the deeper cold water current. Probably not very frequent.” (Penguin)
- “Yes, I see some clouds coming.” (Global Weather)

4 - Claim & Moderate Evidence (provides two pieces of evidence—mentions mechanism and direction, location, or time)

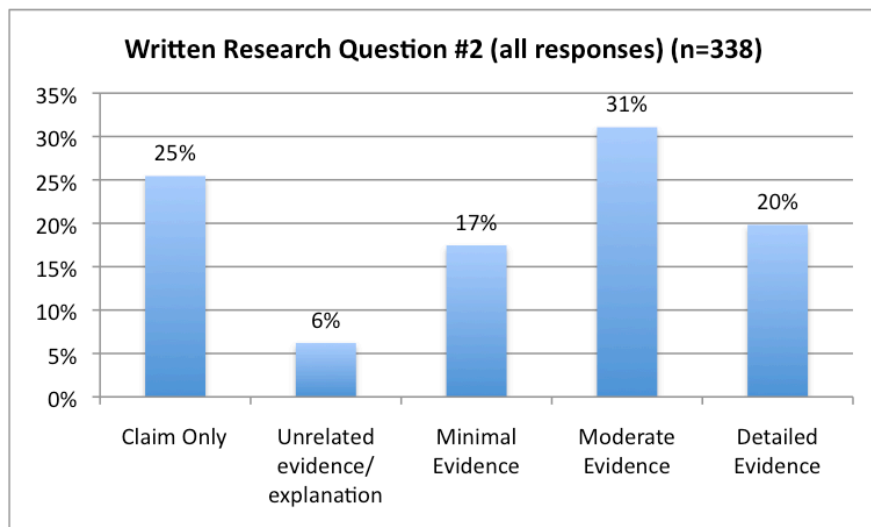
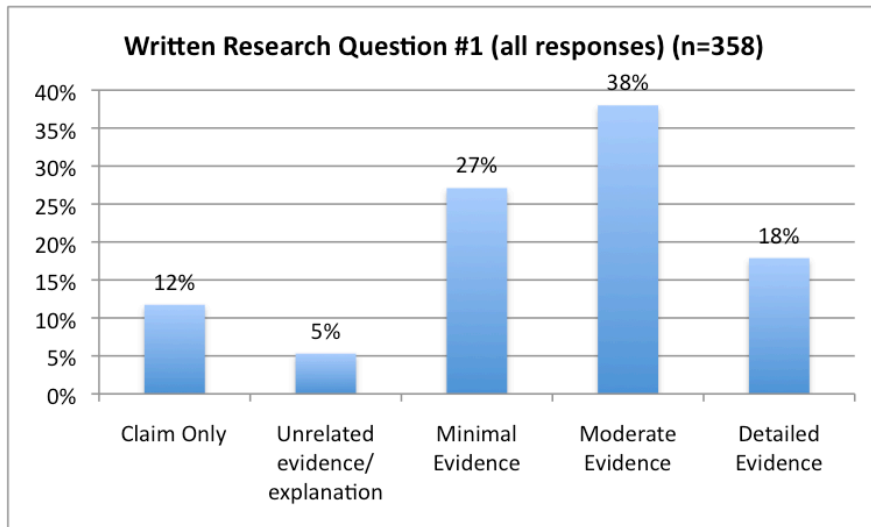
- “Yes, because the “current” arrows are pointing in the direction of the island” (Shoes)
- “Yes because there are currents going from Alaska to California” (Penguin)
- “Yes. It looks like rain will fall in 8-12 hours.” (Global Weather)

5 - Claim & Detailed Evidence (provides specific and detailed information related to mechanism, direction, location, time, etc.)

- “Yes it could. Follow the N. Pacific current around West Coast to the N. EQ. current towards Hawaii.” (Shoes)
- “Yes. Conveyor belt takes feather under water from Antarctica to Alaska. Wind current takes to SF after conveyor belt water warms up and comes to the top.” (Penguin)
- “Yes because the wind pattern indicates that the high clouds will be near Berkeley (7:00)” (Global Weather)

Overall, participants scored higher on the first research question (where they were asked about evidence directly related to the presentation they had seen) than on the second question where they were asked a question using similar evidence but not directly related to the presentation (mean score of 3.45 for question 1 and 3.16 for question 2 ( $p=.001$ )). However, in both cases, participants tended to use at least minimal evidence in supporting their claim. For question 1, 83% of participants used some evidence. And for question 2, 68% used some evidence. Overall, it appears that the programs were effective in engaging visitors in using evidence from the sphere to support their claims.



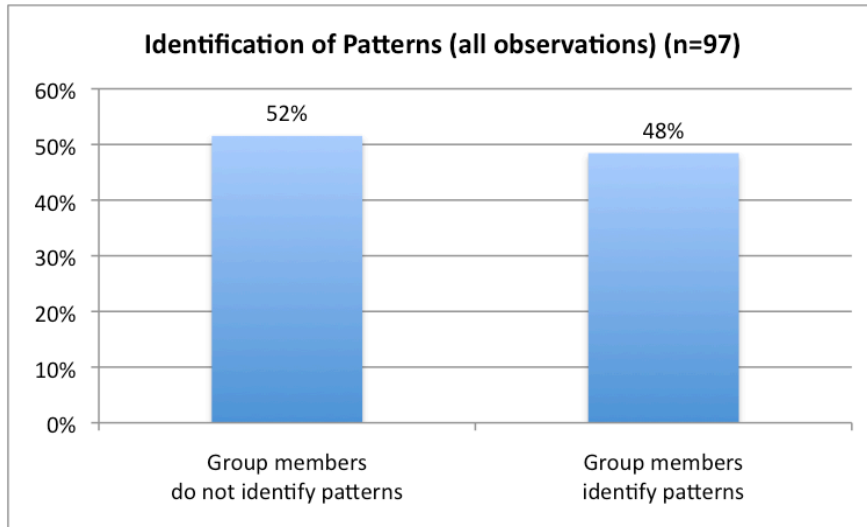


*Finding #2: The LEEGS programs as a group were somewhat less successful at getting visitors to use other patterns of scientific thought (e.g. identification of patterns, presentation of alternative ideas), although they were extremely successful at getting visitors to engage in discussion.*

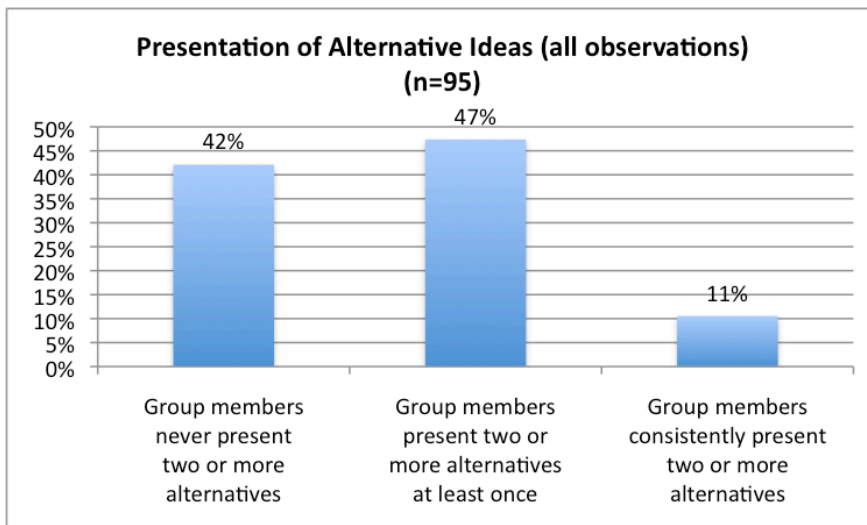
While the primary goal of the program was to engage visitors in using data from the sphere as evidence to support their claims, the project team was also interested in engaging visitors in other patterns of scientific thought. Again, examining all the observations together (regardless of specific presentation or presentation type), it appears that the LEEGS programs were somewhat less successful in engaging visitors in identifying patterns or presenting alternative ideas. However, the program was extremely successful at getting visitors to engage in dialogue with others (particularly within their own group).

Just under half of the observed groups identified patterns during the presentation. For the Go with the Flow presentations, pattern identification typically took place as participants were watching the hands-on demonstration portion of the presentation when visitors identified patterns

in the movement of the water. For the Global Weather presentation, pattern identification again typically took place during the hands-on demonstration when visitors identified air patterns as well as through visitors identifying cloud patterns that could be fronts on the sphere. (Note, as outlined below, some presentation modes were much more effective than others at eliciting this behavior.)

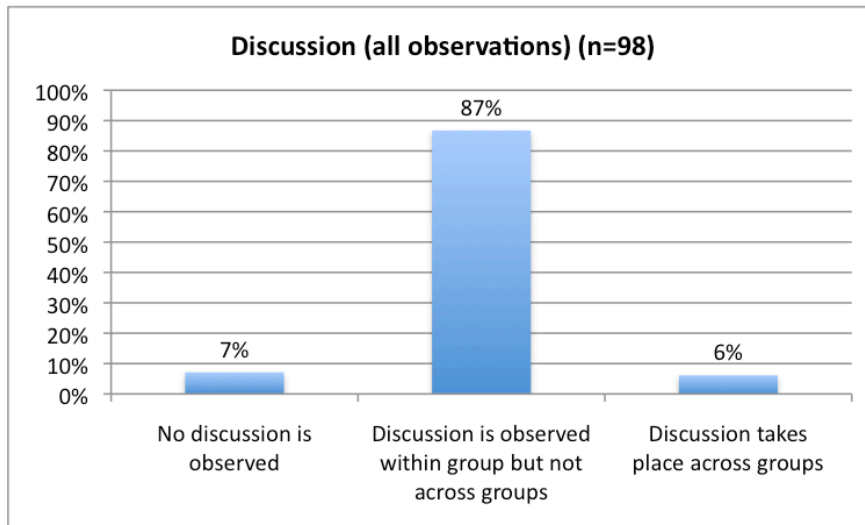


A slightly larger percentage of visitors presented alternative ideas at least once during the presentation (58%). For the Go with the Flow presentations, alternative ideas typically came up when discussing where an object would go and what route it would take to get there. For the Global Weather presentation, visitors most often presented alternative ideas about whether it would rain or not and whether the rain would hit California (or go north). (Again, as outlined below, some presentation modes were much more effective than others at eliciting this behavior.)



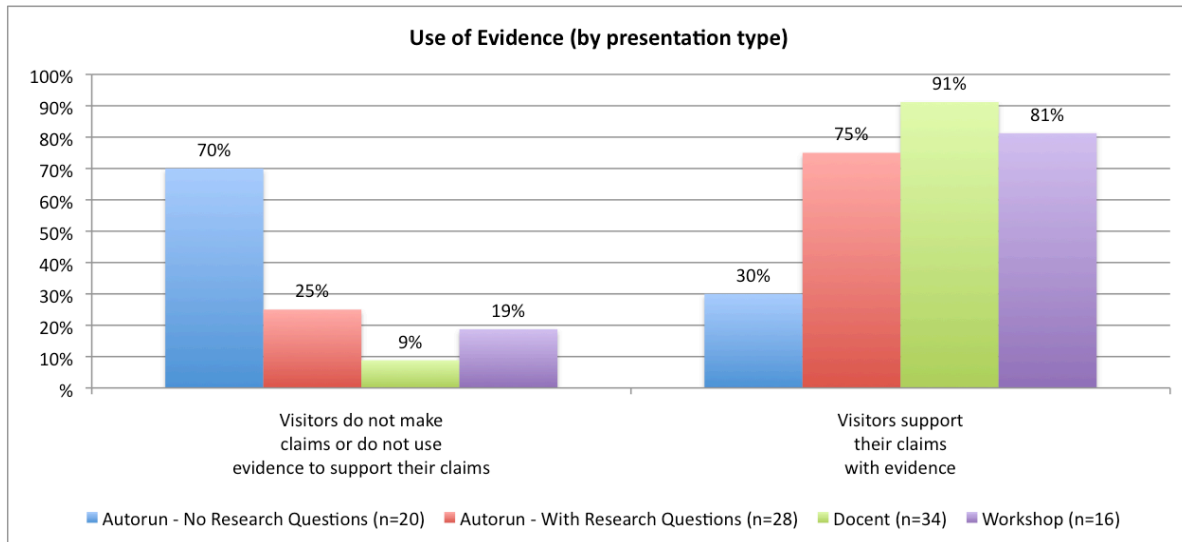
While the presentations were somewhat less successful in eliciting pattern identification and discussion of alternative ideas, they were extremely successful in eliciting discussion in general, particularly among group members within a single group. The program was less successful in

eliciting cross group discussion. When group members did interact with people outside their group, it was with the facilitator, rather than directly with other groups present during the presentation.



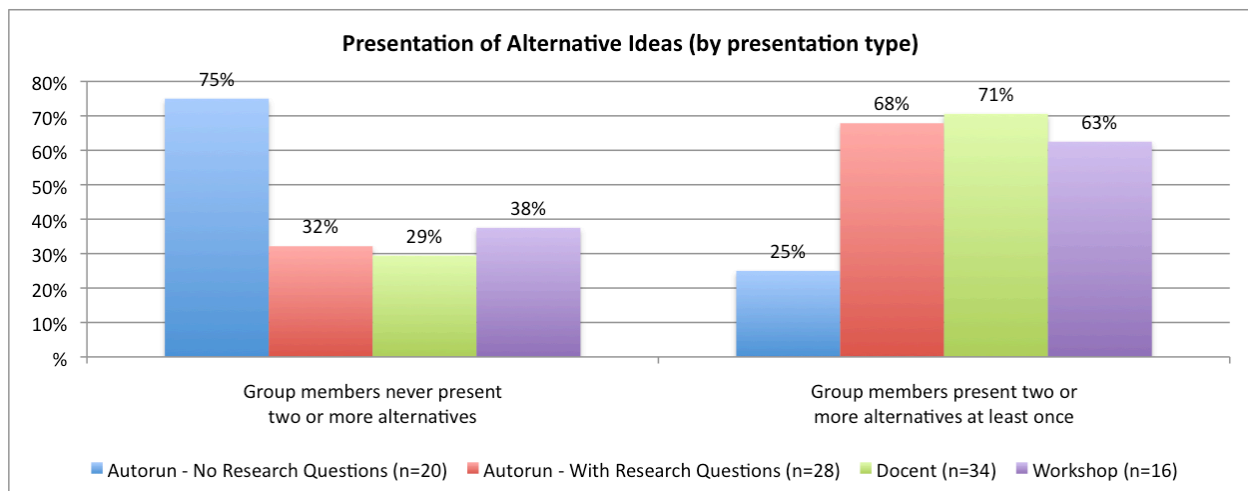
*Finding 3: Workshop and docent-led presentations are more successful than autorun at engaging visitors in verbally expressing scientific thought processes. However, it appears that the use of some scientific thought processes can be increased in the autorun mode simply by stopping the autorun presentation and providing visitors with a chance to work together to express their ideas about what is happening in the presentation.*

Visitors observed during the autorun presentations without research questions (the way the autorun presentations were intended to be run) were significantly less likely than visitors to other presentations to verbally use evidence to support their claims. However, when the research questions were added in to the autorun presentations, the likelihood of using evidence to support their claim increased significantly. There was no significant difference between the autorun presentation with research questions and the docent or workshop presentations. Therefore, it appears that simply stopping the presentation and asking visitors to answer a question can have a significant impact on the use of evidence to support claims.



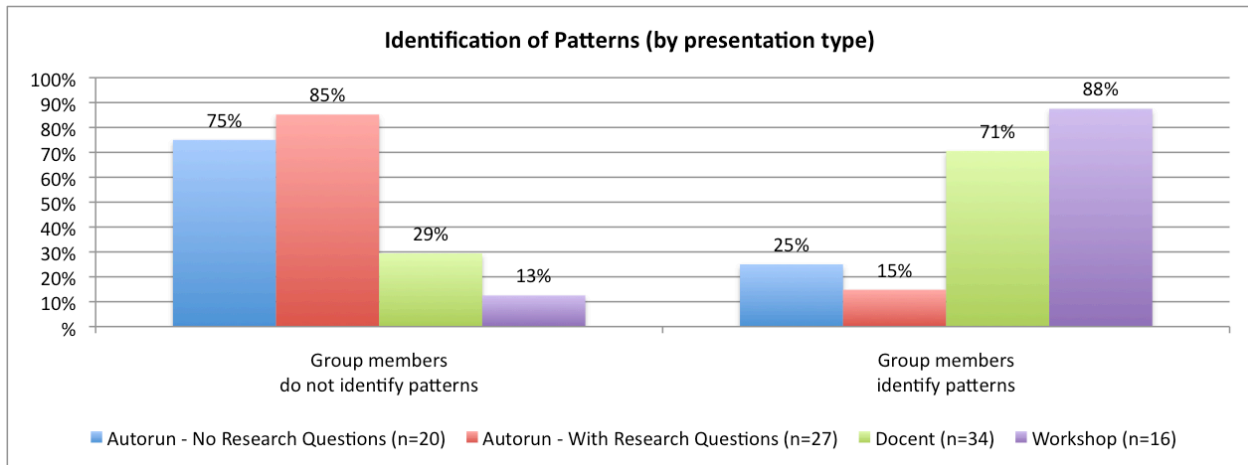
Difference between groups is significant ( $p=.000$ ). Note that categories were collapsed in order to make it possible to run a chi-square analysis.

Similarly, visitors observed during the autorialun without research questions were unlikely to present two or more alternatives. However, again, simply adding the research questions greatly increased the likelihood that this behavior would take place. Again, there was no significant difference between the autorialun presentation with research questions and the docent or workshop presentations.

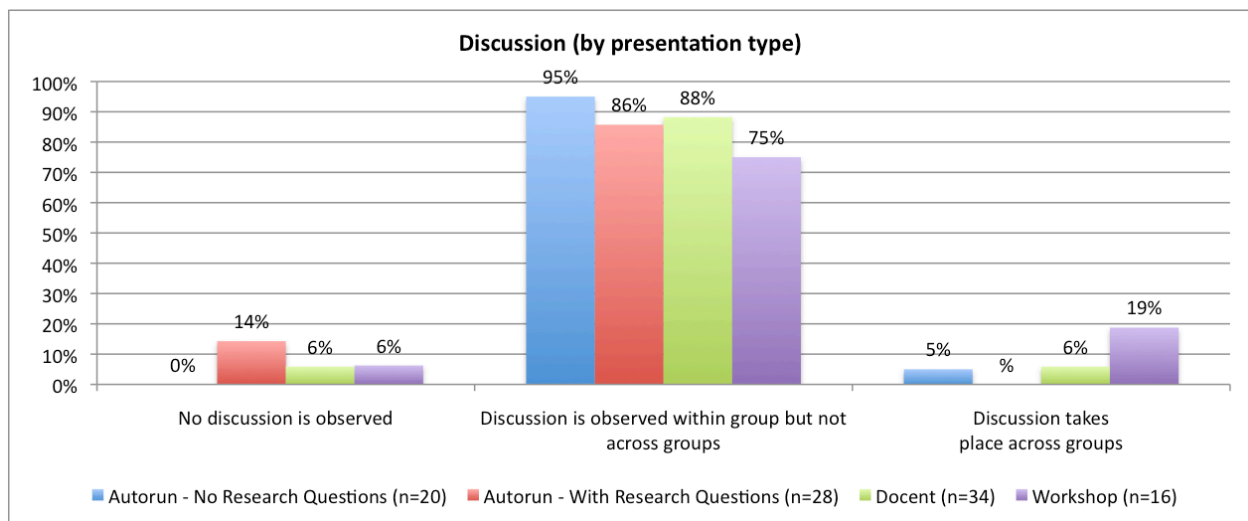


Difference between groups is significant ( $p=.006$ ). Note that categories were collapsed in order to make it possible to run a chi-square analysis.

Identification of patterns did not follow the same trend. In this case, visitors to both types of autorialun presentations were significantly less likely than docent or workshop participants to identify patterns. This is likely because the pattern identification most often took place during the hands-on demonstration. While the autorialun version of the Go with the Flow presentation included video of the hands-on demonstration, it does not appear that the video was as effective at eliciting pattern identification as the live demonstration itself.



Difference between groups is significant ( $p=.000$ )

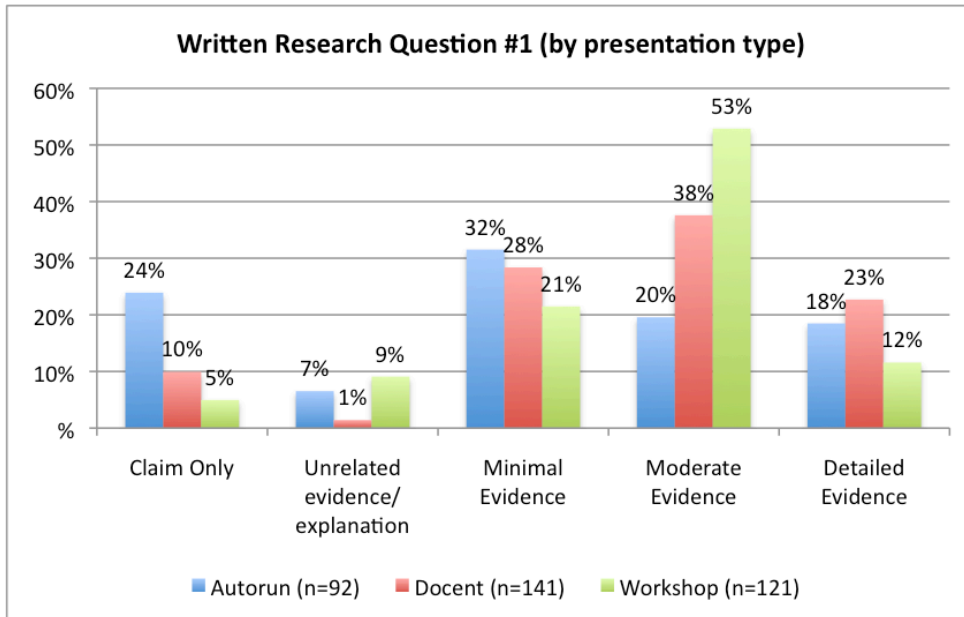


Expected cell counts are too small to run a chi-square analysis.

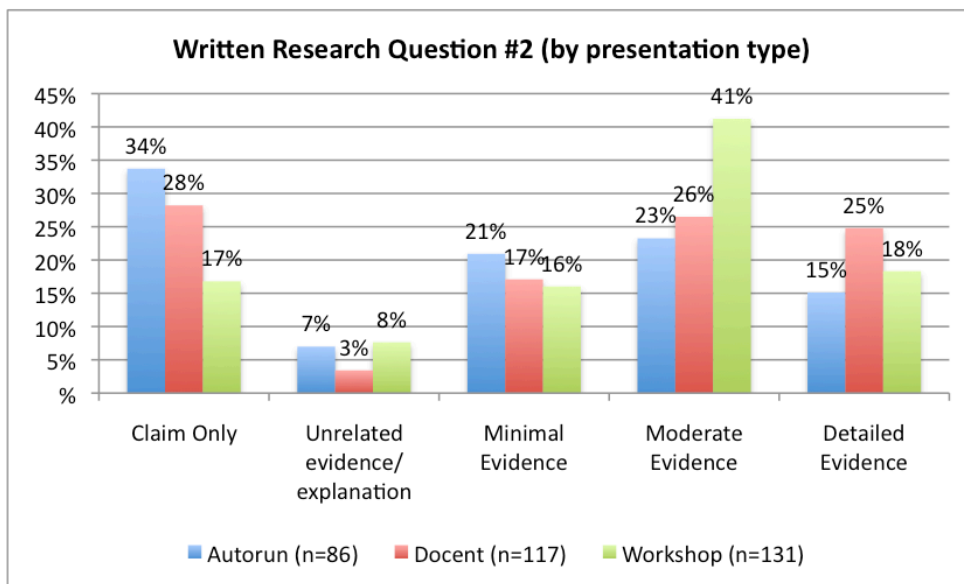
Discussion did not appear to be dramatically affected by presentation type. Even the autorun without research questions elicited discussion in 100% of the observations. This is surprising since autorun presentations might be expected to be watched in silence. However, program developers included a pause in the autorun presentation where they prompted visitors to turn to the person next to them and share their thoughts. In almost every case, this prompt elicited some level of discussion among group members. While it appears that the workshops were most effective at eliciting discussions across groups (which is not surprising since all the groups in a workshop were from the same classroom), there was still surprisingly little cross group discussion during the workshops.

*Finding 4: Workshop and docent-led presentations appear to be more successful than autorun at getting visitors to use moderate or detailed evidence in supporting their explanations. However, this finding does not hold true across all three of the different presentations.*

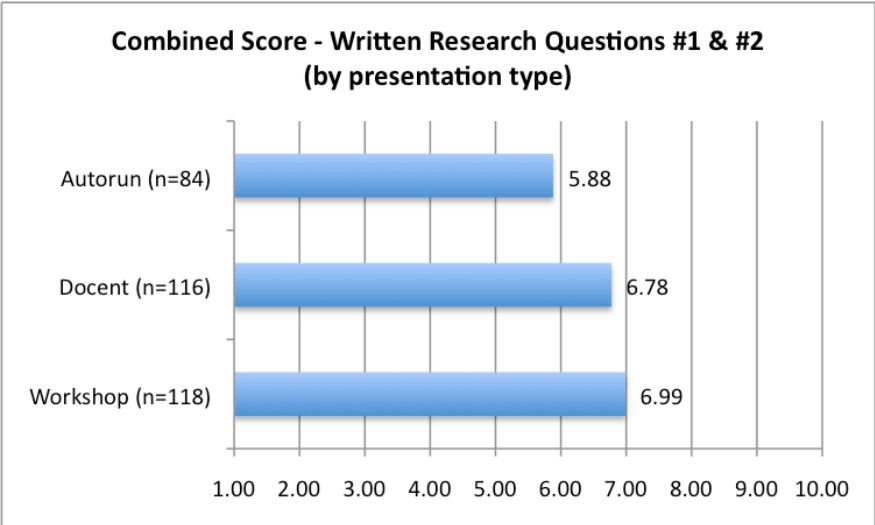
When scores from all the research questions are examined, it appears that the autorun presentation was less effective at eliciting detailed use of evidence. For both research questions, autorun participants were more likely to make a claim only or to use only minimal evidence than participants in other presentations. To compare responses, combined scores from both research questions were tallied and compared by presentation type. Combined scores for responses given during the autorun presentations were significantly lower than scores for the docent or workshop presentations. In both the docent and workshop presentations, visitors are consistently asked to describe what they notice and to think about what evidence might support their claims, which may lead them to more naturally use evidence and to use more detailed evidence when answering the research questions.



Difference between groups is significant ( $p=.000$ )

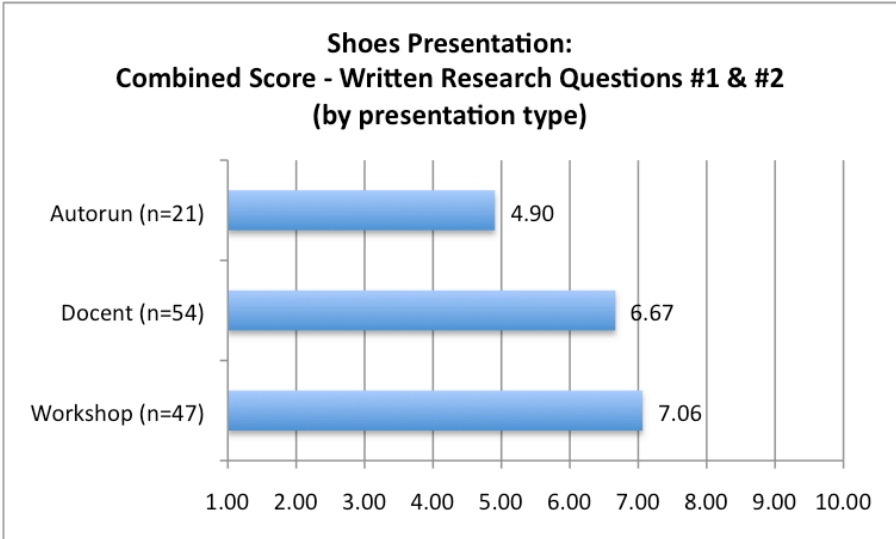


Difference between groups is significant ( $p=.017$ )

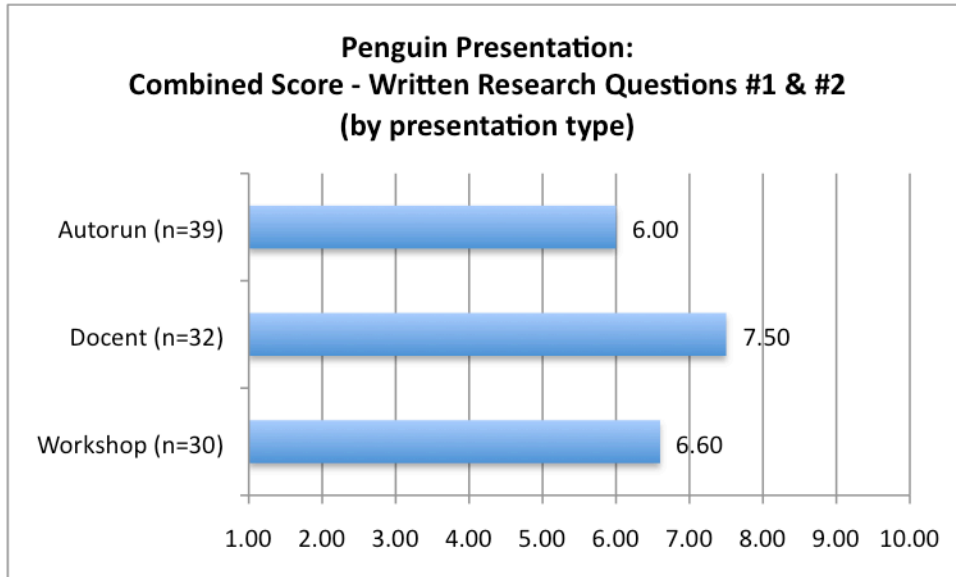


Difference between groups is significant (p=.004)

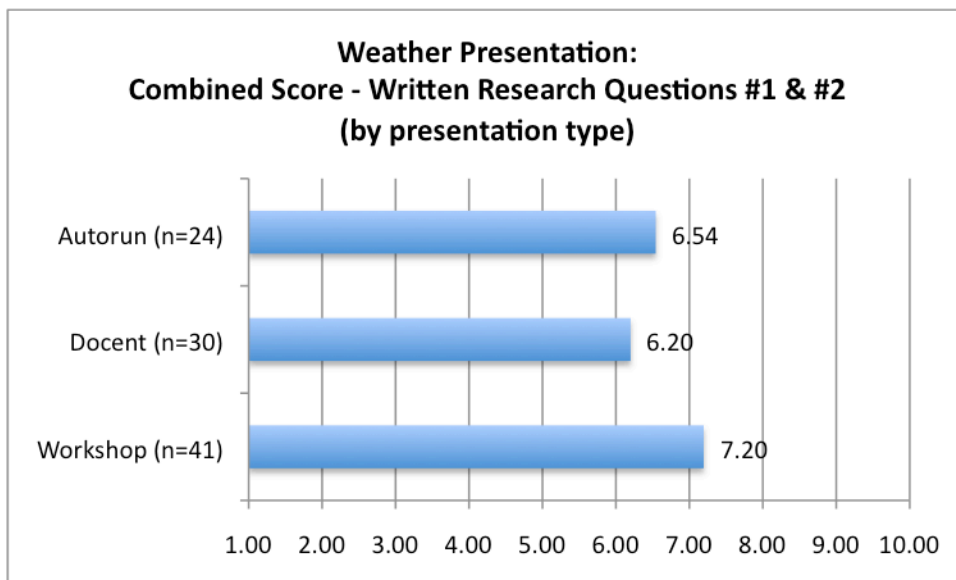
These findings seem to indicate that the autorialun presentations are not as effective at deeply engaging visitors in using evidence to support their claims. However, these findings do not hold true across all the different presentations. For Go with the Flow—Shoes, the combined scores on the research questions are significantly lower for the autorialun presentations than either the docent or workshop presentations. However, for Go with the Flow—Penguin, the combined scores from the autorialun presentations are only significantly lower than the docent presentations (no significant difference between autorialun and workshops). And for Global Weather, there is no significant difference in combined scores between any of the three different presentation modes. The scores for the Go with the Flow—Shoes autorialun were particularly low compared to the other modes and to the autorialun scores for the other presentations. Additional data on the differences between the Shoes autorialun and the other Shoes presentations and between the Shoes autorialun and the other autorialun presentations would need to be collected to fully understand why these scores were so much lower than the scores for the other presentations.



Difference between groups is significant (p=.001)



Difference between groups is significant ( $p=.016$ )



No significant difference between groups.

Because the program developers were less interested in content knowledge development and more interested in engaging visitors in using scientific thought processes, we did not code the written responses for content understanding. However, there was a somewhat common confusion that became apparent in looking at the responses. For both Go with the Flow presentations, there appeared to be some confusion about whether the objects (shoes, feather) were being pushed by currents or the wind itself. Some of the responses seemed to indicate that participants thought the objects were being moved by the wind itself. Examples included:

- Yes because which ever the wind is blowing the shoes go where the wind is blowing (Shoes)
- We think it could go to California because the wind is blowing from Antarctica to California (Penguin)

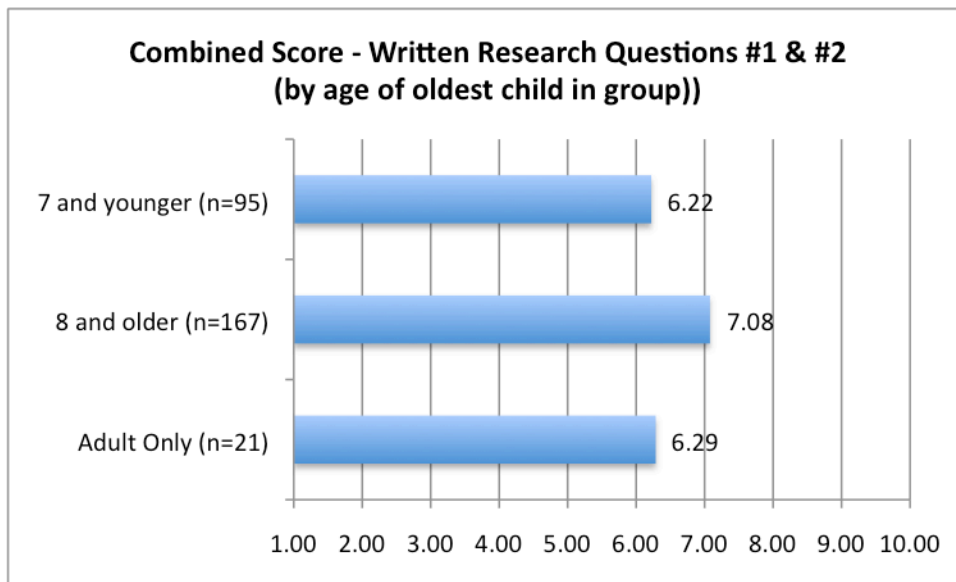


- Follow current to Calif., then wind will blow to shore (Penguin)
- Yes, because the feather could get into the Jet current and fly to California (Penguin)

For the Shoes presentation, 13% of respondents had answers that indicated that the wind was moving the shoes for question 1 and 9% for question 2. For the Penguin presentation, 15% responded this way for question 1 and 11% for question 2. Additionally, for both presentations, many additional respondents mentioned “wind currents” as the mechanism moving the object. It was unclear whether these respondents also thought the wind itself was moving the object or whether they were actually referring to wind-driven currents.

*Finding 5: Groups where all the children in the group are 8 or older are more successful than groups with younger children at using moderate or detailed evidence in supporting their explanations.*

In terms of the verbally expressed patterns of scientific thought (e.g. identification of patterns, use of evidence to support claims, presentation of alternative ideas), there was no significant difference based on age of the children in the group. However, for the written research items, there was a significant difference based on the age of the children. Groups that included children older than the age of 8 received significantly higher combined scores on their research questions than groups that only included children younger than 7. There was no significant difference between groups with children age 8 and older and adult-only groups. This indicates that the program may be best suited for engaging children age 8 and older and some modifications may be necessary for engaging younger children in using evidence from the sphere to support their explanations.



## RECOMMENDATIONS

There are a number of findings in this report that could be used to guide the development of future presentations for the Science on a Sphere system. The following recommendations are based on the findings of this evaluation report.

- **Continue to develop presentations that engage visitors in making predictions based on evidence:** Developing Science on a Sphere presentations that focus on presenting data

and asking visitors to use that data to make a prediction appears to be an effective way of engaging visitors in using scientific thought processes, especially using evidence to support their claims, but also, to a lesser degree, identifying patterns and presenting alternative ideas. Future presentations that use this approach will be a valuable tool in continuing to promote these behaviors among museum visitors.

- **Include prompts and pauses in autorun programs to allow visitors to discuss ideas:** The use of a simple prompt asking visitors to share ideas with one another and a pause in the autorun presentation was surprisingly effective at getting visitors to engage in discussion with one another. During all of the autorun observations without research questions, groups were observed engaging in discussion with one another (at least minimally). This simple technique can be implemented in future presentations to engage visitors in discussing ideas.
- **Think about ways all the presentations can be designed to increase cross-group discussion:** While the presentations were very effective at getting visitors to talk to one another within their group, there was very little discussion between different groups. During the docent-led interaction, groups primarily interacted with one another or with the docent. Even during the workshop, where all the participants were from the same school, participants tended to only speak with the small group of other students they were working with. Thinking about ways presentations can be designed to elicit more sharing of ideas between groups would be valuable.
- **Design the autorun programs to include opportunities for visitors to share ideas with one another:** The autorun programs were much more effective at engaging visitors in using scientific thought processes when the research questions were used, that is, the presentation was stopped and group members were asked to work together to respond to a question. Thinking about ways to incorporate this type of activity into all the autorun presentations would make them significantly more effective at achieving their goal.
- **Design the autorun programs to include opportunities for pattern identification:** Because pattern identification typically took place during the hands-on demonstration portion of the presentations, this behavior did not occur nearly as frequently in the autorun presentations. The videos of the hands-on demonstrations did not appear to elicit the same types of pattern identification behavior as the live versions of the same demonstrations. If pattern identification is a goal of the program, it would be good to think about modifications that could be made to the autorun presentations to encourage this behavior.
- **Think about ways to better engage families with young children in the presentations:** The presentations worked best at engaging children over the age of 8 in using evidence to support their explanations. Additionally, although this was not tracked as part of the evaluation, anecdotally researchers noticed many families with young children leaving the presentation early. Thinking about ways the presentation could be modified to engage younger children may be useful especially for those formats designed for family groups (autorun and docent-led).