

Exploring the Universe from Antarctica - An Informal STEM Polar Research Exhibit

Ross Tredinnick¹, Rebecca Cors², James Madsen³, David Gagnon⁴, Silvia Bravo-Gallart³, Bryce Sprecher¹, and Kevin Ponto¹

¹Wisconsin Institute for Discovery, University of Wisconsin-Madison; ²Wisconsin Center for Education Research, University of Wisconsin-Madison; ³Wisconsin IceCube Particle Astrophysics Center, University of Wisconsin-Madison; and ⁴Field Day Lab, Wisconsin Center for Education Research, University of Wisconsin-Madison

Keywords: Informal Science Education, Virtual Reality, Touch Table, Concept Map, Museum Exhibit

Publication Date: April 20, 2020

DOI: <https://doi.org/10.15695/jstem/v3i1.03>

ABSTRACT: Many concepts in astrophysics research can be difficult for a lay individual to understand or to comprehend their importance. One such example concept is the IceCube Neutrino Observatory, which detects high-energy neutrinos at the South Pole in Antarctica. The observatory uses information from detected neutrinos originating deep in outer space to better understand astrophysical phenomena like black holes or exploding stars. Unfortunately, it is often difficult for the public to understand how these pieces fit together towards creating a more complete understanding of our universe. To promote public understanding of IceCube, an interactive exhibit was created which used large multi-touch screen and Virtual Reality (VR) equipment. The exhibit, placed in a public environment, was evaluated both formally and informally around its effectiveness of providing STEM learning opportunities. The results show that the system not only provided an effective means of conveying content, but also provided a means of sparking the curiosity of users to learn more about the presented subject matter. Further results show that outcomes demonstrated significant differences between subject responses depending upon which of the two deployed devices they used. Findings also provide evidence that retrospective survey designs have the same rigor in data collection as traditional pre-and posttest designs when investigating exhibits. Based on these findings, guidelines are offered for others who aim to deploy similar systems in publicly accessible spaces.

INTRODUCTION

Many concepts in astrophysics research can be difficult for a lay individual to understand or to comprehend their importance. One such example concept comes from the IceCube neutrino observatory, which studies high energy neutrinos (IceCube Collaboration, 2013; IceCube, 2018). Neutrinos are microscopic particles that originate light years away, as a result of cosmic phenomena, such as exploding stars and black holes. Relating the many concepts surrounding neutrinos and their detection to the general public has been a long-standing area of focus and concern from members of the Wisconsin IceCube Particle Astrophysics Center (WIPAC), an organization heavily invested in the IceCube research project. WIPAC has worked together with other organizations to produce a feature film, and regularly participates in University sponsored youth camps. Creating a museum exhibit was a logical continuation of the polar research team's outreach efforts.

Recent advances in low-cost VR technology have enabled new directions in informal teaching and learning. In

the recent past, head mounted display (HMD) technology cost upwards of twenty thousand dollars for a HMD. Today, the same hardware costs between fifty to fifteen hundred dollars, depending on the particular type of HMD. This cost reduction makes it much easier for informal learning centers such as libraries and museums to adopt HMD VR technology. Challenges for these informal learning centers have shifted from procuring the hardware to the development of engaging content (Stogner, 2011)

One advantage that VR provides is its ability to place users within a 3D environment that otherwise would be difficult or impossible to inhabit. This fact bodes well in efforts towards educating the public on areas on Earth that are difficult to reach, such as the Polar regions of the Arctic Circle and Antarctica. The following research disseminates results of a study that compares two informal STEM learning museum style experiences, one employing VR technology, and one employing a multi-touch high resolution large display screen, also known as touch table (TT) technology. The ex-

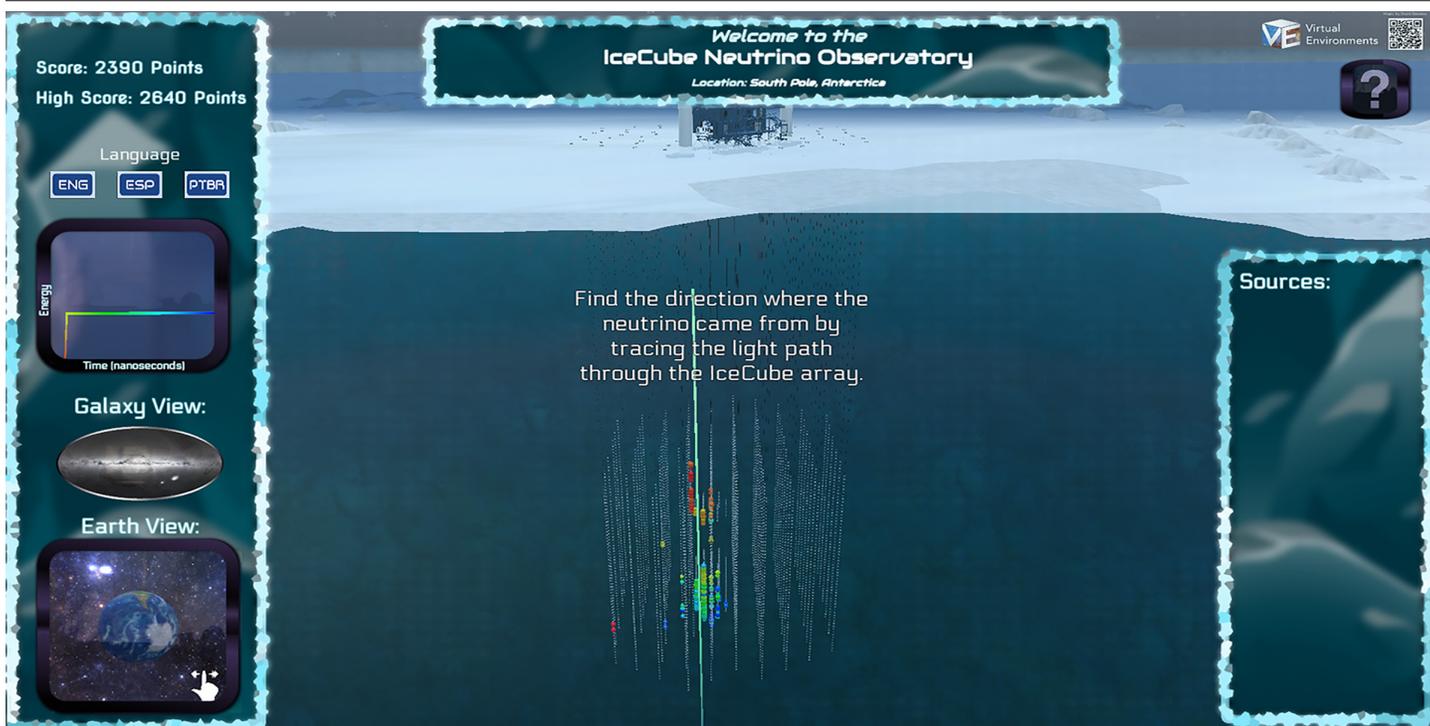


Figure 1. Screenshot of the touch table exhibit during game play. The user's viewpoint slowly rotates around the central array and Antarctic environment. Multiple users can swipe their fingers across the array to try and detect directions of neutrino events and learn of what sources created the neutrinos (kept track of in the list on the right). Features on the left side include, changing language (English, Spanish or Portuguese), viewing a graph of the energy over time of the current neutrino event, and views of where a neutrino flew through earth as well as directionality on a galaxy map.

hibit aims to teach users about the fundamentals of a large-scale research project taking place at the South Pole in Antarctica.

To create this exhibit, WIPAC worked with members of a VR research group, the Wisconsin Institute for Discovery (WID) Virtual Environments group, and an educational games research group, the Field Day Lab, both located on the University of Wisconsin – Madison campus. The collective goals of the effort were to create an exhibit that contained two different interactive experiences about IceCube research and to compare informal learning outcomes between the two experiences. The team deployed the exhibit in a publicly accessible main floor of the WID. The team chose to deploy the exhibit here, rather than a local museum or library, due to a partnership with the building's public outreach team, "Discovery Outreach", as specified and planned through the project's funding source (NSF). A mutual interest between the research team and Discovery Outreach existed to bring new, interactive content to the publicly accessible main floor of the building. The research team gathered results through developed surveys during a three-and-a-half-month long time period. This project seeks to answer the overarching question: can interacting with the exhibit promote public understanding of complex, high-tech scientific research like IceCube? The following section highlights work related to employing TT technology and VR technology in a museum context. After reviewing related work, the next section out-

lines how the team developed the exhibit. Following this, the document reports findings both on answers to the overarching question, as well as on additional hypotheses regarding the exhibit. The report will finish with a discussion of results, limitations, and future directions.

RELATED WORK

Prior efforts exist that incorporated VR components into museum exhibits (Grafe et al., 2002), (Lepouras and Vassilakis, 2004; Wojciechowski et al., 2004; Tost and Economou, 2007; Hirose, 2015). The previous work most closely related to IceCube occurred in the development of an astrophysics VR museum exhibit (Podgorny, 2004). This work sought to incorporate gamma-ray telescope data together with Sloan Digital Sky Survey telescope data into a VR museum exhibit with a target audience of elementary school children and families (York et al., 2000).

Common challenges for museum exhibits delve into the field of user experience design (Hornecker and Stifter, 2006). One common challenge users have is the need to adjust to the context of the informal learning situation, which has been studied by informal science researchers since the 1970s. Falk and colleagues (Falk et al., 1978) described the 'novelty field-trip phenomenon' as "an adjustment and adaptation process that learners use in response to initial feelings of disorientation when they arrive at out-of-school learning

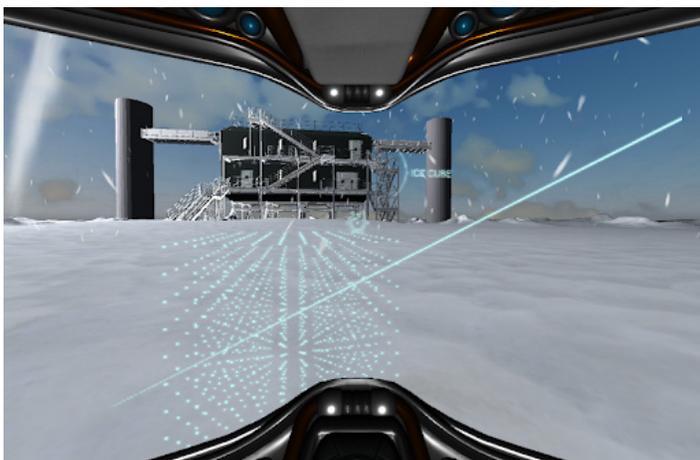


Figure 2. Screenshot of the VR exhibit. This is the first person view of a user near the beginning of the experience where they are situated on Antarctica near the IceCube lab research building.

place, whose setting is typically designed to be stimulating” (Cors, 2016, p. 30). Researchers have explored ways to orient users by providing an intuitive means to let the user, whom may have no prior knowledge about neither the technology nor subject matter, know how they should proceed through an exhibit. This holds particularly true for VR-based exhibits that employ HMD technology due to the fact that an HMD blocks a person’s view of the real world. In addition, audio narration, music, or sound effects may be playing through the HMD, so a user may not be able to hear sounds from the real world. This means that users may not be able to easily rely on assistance from docents, or if they can, it may not be easy for the user and the docent to communicate effectively with each other. One option to alleviate this is to build a VR museum experience to progress through the use of narrative (Roussou, 2001). The VR exhibit for this project, makes use of narrative to both guide a user through the experience and to provide educational content pertaining to neutrinos and Icecube.

Roughly a decade ago, prior to the reduction in costs for VR technology, the popularity of adopting multi-touch tables for museum exhibits reached a high level (Correia et al., 2010). Most prior work has focused on evaluating the effectiveness of the technology in terms of ease of use and overall “success” of the TT exhibit (Hornecker, 2008; Ciocca et al., 2012; Horn et al., 2012; Van Dijk et al., 2012). Other work has analyzed TT exhibit impact on learning outcomes (Block et al., 2012; Zaharias et al., 2013). Zaharias et al. (2013) compared 5th grade student learning outcomes between traditional printed maps versus similar content presented on an interactive TT. Their work found no significant differences between learning outcomes, but found high levels of user experience. Block et al. (2012) succeeded in creating an interactive information visualization based museum exhibit for exploring the tree of life, and found that the exhibit facilitated learning across multiple age levels, while also succeeding

in attaining a desired Active Prolonged Engagement (APE) with its users (Humphrey and Gutwill, 2017).

The presented work creates a combined exhibit that co-locates the pre-existing popular TT museum exhibit technology with VR, a technology that has increased vastly in its accessibility over the past five years. This combined setup offered a unique opportunity to understand people’s basic learning, curiosity, and engagement across each experience. The following section provides background details on the exhibit and their development process to better understand the subsequently presented research study and findings.

EXHIBIT DEVELOPMENT

To create the exhibit, members of the research team employed a popular 3D game development engine, Unity (<https://unity.com>). Unity has become a standard development platform for design and creation of games and VR experiences. Prior to implementation, the research team paired an iterative design methodology together with several interviews of members of WIPAC to better understand concepts related to astrophysical neutrino research and their research at the South Pole. The team used an Oculus Rift CV1 HMD for development of the VR experience, and used a 65” 4K resolution 10-point multi-touch screen for the development of the TT experience.

One goal of the project involved creating experiences that lasted one to three minutes thus allowing a larger number of people to partake in the exhibit. Initially, the team considered developing a single, shared 3D environment that users would experience simultaneously regardless of which type of hardware a person used, through networking the two experiences together. The one- to three-minute goal, considered to be of greater importance in a public space, led the team away from the shared, networked, environment due to various complications of such a setup. For example, if two people are not both interacting with the exhibit at the same time, the exhibit would still have to seamlessly handle a single user for either experience. Instead, the team decided to share visual and audio assets between the two experiences, which provides a common setting and feeling of them being together, even though each experience and interface is different. A screenshot of the TT experience is shown in Figure 1, while a screenshot of the first person view of the VR experience is shown in Figure 2.

Touch Table Experience. The team designed the TT experience with the goal of creating a multi-touch game that is fun to play and that teaches users a basic understanding of neutrinos and IceCube research. Through iterative design of the TT experience, the final version took the form of a fast-paced puzzle game. In the game, any number of users can approach the TT and interact with it via finger presses and

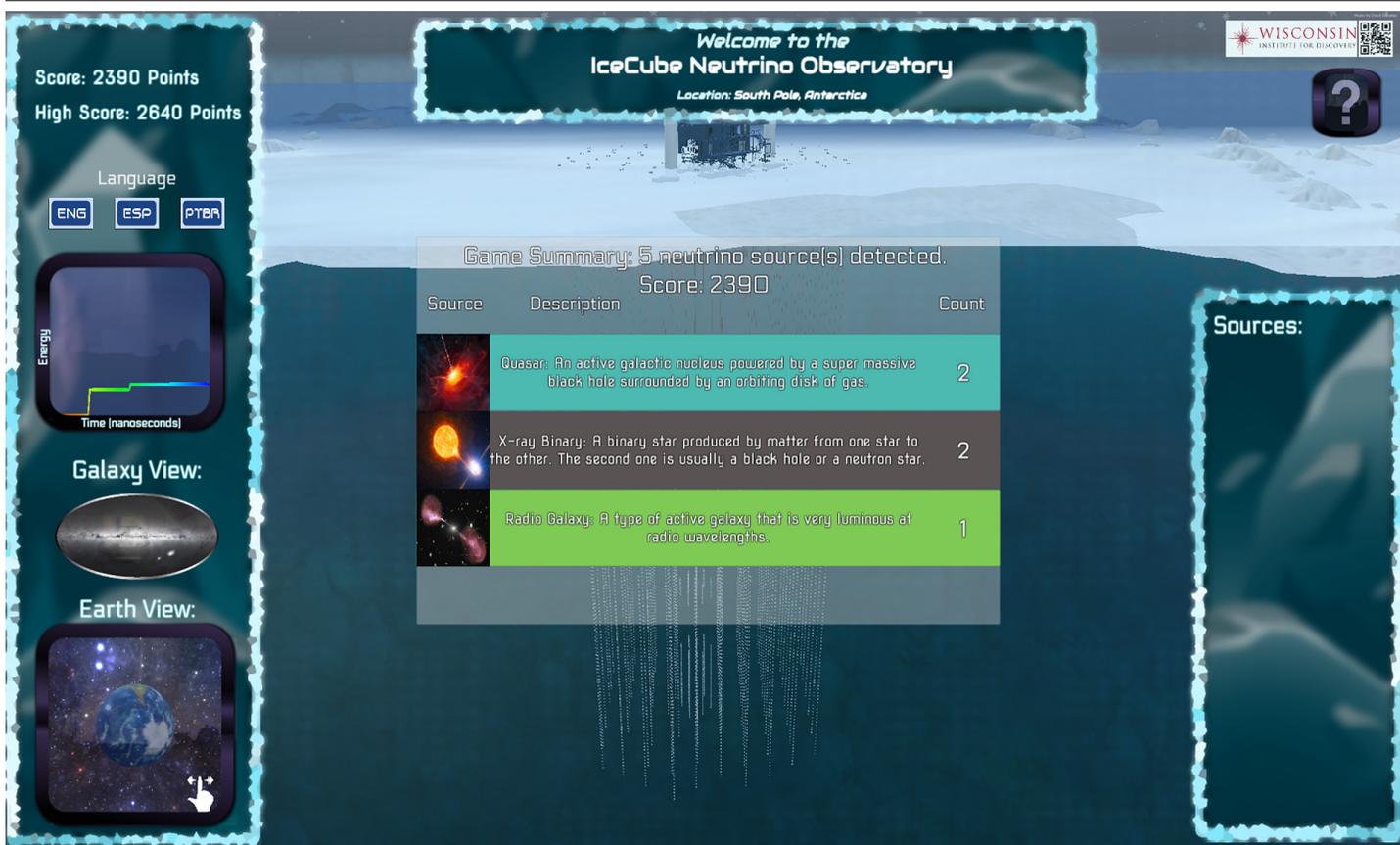


Figure 3. Screenshot of the Touch Table experience summary screen at the end of playing a game session. Basic information is presented to the user about cosmic phenomena that act as the source of the detected neutrino events.

swipes. The team developed the TT experience using Unity and a collection of 3D models provided by the WIPAC or obtained through 3D model repositories, such as creative commons. The player takes the role of an IceCube scientist whose goal is to determine the direction of neutrino events as they propagate through the IceCube array. The primary view of the game shows the IceCube laboratory on top of the South Pole's ice sheet, with the IceCube detector array visible underneath the ice. The detector array is made up of over 5,000 digital optical modules (DOMs) frozen in the ice below the surface. This array indirectly detects when a neutrino has passed through it. When neutrinos interact with water molecules in the ice, a blue light is emitted, this is known as Cherenkov radiation. This light is detected by the DOMs and a signal is sent up cables to the IceCube lab. When a series of DOMs detect light in a close time sequence this is known as a neutrino event. An event signifies that a high energy neutrino may have passed through the array. Therefore, neutrino events are very important to IceCube scientists as the events serve as the signals for when neutrinos have passed through the detector. Upon receiving the signal of a neutrino event, the next step in the researcher's process is to understand what direction the neutrino event passed through the array. Understanding a neutrino event's direction allows astrophysics research labs to focus their analysis on the same direction and same timing to see if they detect any astro-

physical phenomena. It is this combination of data analysis, or multi-messenger astronomy, which can verify the astrophysical phenomena (IceCube, 2018).

The game represents the neutrino events through a visualization method used by the IceCube scientists. A sequence of spheres is drawn on each DOM that detected light, and sphere radius depends on the amount of light detected. Spheres are time coded on a rainbow color scale, with red representing time early in the neutrino event and purple representing a later time (Figure 1). The game shows the neutrino events on a much slower time scale than in real life; real life neutrino events occur on nanosecond time-scales. When the events appear within the array, the player must use the visualization to determine the direction they think the neutrino is coming from, and swipe through the array in that direction. If the user's initial swipe is within 95% of the neutrino event's direction a sequence of three additional panels appear. In these three panels the user refines their directional estimate by swiping on three additional views of the event: one top down view, one front view, and one side view. This interaction paradigm follows from prior outreach activities that WIPAC conducted for informal learning of neutrinos. As a user swipes all three panels the game re-calculates the accuracy of their neutrino detection. The target accuracy starts at 60% and increases or decreases depending on how well the user is playing. If the user achieves the target ac-

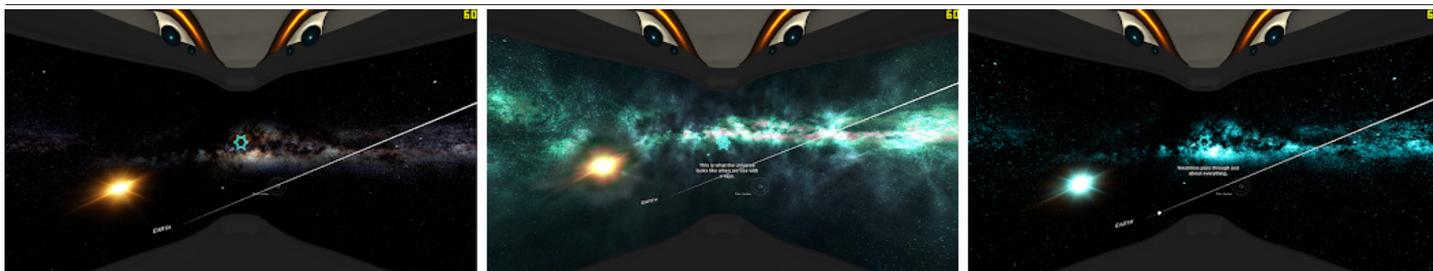


Figure 4. Three views within the VR exhibit of different wavelengths, visible (left), x-ray (middle), neutrino (right)

curacy, they discover the source of that neutrino event. The sources are one of eight possible cosmic phenomena known to produce neutrinos, such as blazars, quasars, or black holes (Mohapatra and Pal, 2004). After sixty seconds, the game shows a summary screen of detected sources along with an overall score. The summary screen contains descriptions of each cosmic phenomena so that users can learn about the sources of the neutrino events that they detected. Figure 3 shows an example summary screen. The summary screen displays for an additional sixty seconds before the game restarts to the tutorial.

The Virtual Reality Experience. The VR experience takes users on an immersive journey from the IceCube lab at the South Pole (Figure 2) through four different levels of outer space, ending in deep space where they are face-to-face with a black hole. Progress in the VR scenario occurs via gaze-based navigation where the user follows the directions provided by a narrator. The experience provides both audio cues and captions for all presented narration. At the beginning of the scenario, the user performs orienting tasks, such as looking at targets within the environment, and receives introductory information. Next, the narrator gives background information on neutrinos and IceCube, before the experience shows a neutrino passing through the DOM detector array. The direction of the neutrino is shown as a line segment with a target at its end. Upon gazing at this target, the user is warped through a portal into the solar system near Pluto. At this location, additional background information is given on neutrinos and the user is introduced to a gameplay mechanic where they can switch between visible light, x-ray, or neutrino vision. Each view allows the user to see the surrounding world from the corresponding visible spectrum. Figure 4 shows examples of each view. After switching between the views, the user travels through another portal to the outer limits of our galaxy. The narrator provides additional information about how the user will come face-to-face with some cosmic phenomena that is thought to be the source of the neutrino that was detected by the IceCube observatory at the beginning of the voyage. The narrator explains to the user that they must switch between the three views and gaze at the cosmic phenomena in order to identify and download information from the neutrino source. Traveling through the

last portal brings the user face-to-face with an artist's rendition of a black hole. After the user switches between the three views and gazes at the black hole, the user can travel back to earth and hear the final information about IceCube, their journey, and neutrinos.

RESEARCH METHODS

The final exhibit setup used during the study is shown in Figure 5. To collect data regarding informal STEM learning, the team developed Qualtrics survey instruments and incorporated them into the exhibits (Qualtrics, 2014). Two



Figure 5. The final exhibit installation as situated in the public research institute main floor. The exhibit was situated within and around the physical structure known as the “3D niche” that houses additional outreach exhibits for campus research projects.

Table 1. Variable Operationalization Table considering factors the team desired to learn through the exhibit, type of data gathered, draft of survey item, and which survey item was used in. IV=Independent Variable; DV=Dependent Variable, N=Nominal (AKA Categorical), O=Ordinal; I=Interval; R=Ratio; P=Open Ended

Factor	Type	Survey Item (Source)	Post	Pre-Post
Person	IV			
Age	I	What is your age?	X	
Gender	N	What is your gender?	X	
Experience	IV			
# VR and # TT plays	R	Specify how many times you played each exhibit:	X	
Usability	O	I found the VR / TT exhibit easy to use.	X	
Usability	P	Help us by describing any difficulties, if any, you had with the exhibit.	X	
Take-Aways	DV			
Basic Understanding - IceCube	N	Where is IceCube located?	X	X
Basic Understanding - IceCube	N	Touch and Drag 5 words that have to do with IceCube into the Box (Concept Map)	X	X
Basic Understanding - Neutrinos	N	Where do neutrinos form?	X	X
Basic Understanding - Neutrinos	N	Touch and Drag 5 words that have to do with Neutrinos into the Box (Concept Map)	X	X
Gain in Basic Understanding	O	Rate your understanding of IceCube before the exhibit / Rate your understanding of IceCube after the exhibit	X	X
Curiosity	O	After seeing the exhibit, I would like to know more about IceCube	X	
Attitude	O	The research work at IceCube is important for society.	X	

versions of the survey were made to accommodate two different research samples: a main post-sample (N=154) and a pre-post sample (N=31). The post-sample version of the survey was completed after subjects experienced either the TT and or the VR experience (or both). Each survey had the same questions, but different programming was required to build one survey into the TT and the other survey into a tablet in a stand next to the exhibit. This allowed for users to optionally complete either survey, regardless of whether they played one or both experiences (each survey asks them how many times they played each experience). Users accessed the TT survey via a pop-up dialog that appeared at the end of the game that asked if they would like to fill out a survey. If yes, the survey appeared in a pop-up web browser. The browser automatically closed after the user finished the survey, or if the survey timed out (to handle the case where a user leaves mid-way through the survey). The post-survey took about two minutes to complete. When a minor filled out the survey, it included a parental consent checkbox that had to be checked prior to starting. All instruments and research procedures were approved by the University of Wisconsin - Madison Institutional Review Board.

When developing the surveys, the team considered factors important to a number of project stakeholders for steering of question development and reviewed example survey items used by other informal learning researchers. Stakeholders included: the research team, NSF, the public, and the IceCube research team. Table 1 shows a variable operationalization table and associated survey questions for each variable. Each survey went through an extensive testing phase

for two weeks. Test data was collected and evaluated by the team to determine correctness and to observe any deficiencies in the survey setup or data collection process.

Then, the team collected post-sample responses between March 30th, 2018 and July 18, 2018. A total of 154 people completed post-sample surveys. Of those 154 people, 75 took the survey on the tablet next to the VR and 79 took the survey embedded in the TT. In the 154-participant group, 89 were female and 50 were males, and 15 preferred another gender identification. The age demographics included 75 participants between the ages of 18-30, 30 between the ages of 31-45, 32 between the age ranges of 46 and 60, and 17 aged 61 or older.

In addition to the post-sample, the team also administered a pre-post sample survey on July 26th, 2018 to 31 people. The pre-post sample consisted of 19 females, 12 males, with 80% of participants between 18-30 years old. Amongst the pre-post sample, 11 people tried the VR experience, and 20 tried the TT.

The pre-post sample survey questions are indicated on Table 1. Each pre-post sample survey took approximately one minute to complete. Two members of the research team recruited pre-post participants by soliciting from passers-by within the main floor of the research institute where the team had installed the exhibits. Persons were asked whether they wanted to participate in a research study involving informal STEM learning and VR, and consent occurred before the first question of the pre-survey. The pre-post sample served as a cross-validation of the post-sample along with providing additional information on gain in basic understanding.

Table 2. *Perceived Basic Understanding Question Results for each the tablet/VR and TT surveys. Post Survey, Pre-Post Survey, and One-Play results are shown.*

Factor	Test	Mean Before Game (M_B)	Mean After Game (M_A)	Gain	p	t	η^2
TT Post Survey Basic Understanding N=79	Paired Samples t-test	1.9 (SD=1.6)	3.3 (SD=1.6)	1.4	< 0.001	t(78) = -9.11	0.52
VR/Tablet Post Survey Basic Understanding N=75	Paired Samples t-test	1.8 (SD=1.2)	3.9 (SD=1.2)	2.1	0.002	t(74) = -12.98	0.69
Pre-Post TT Basic Understanding N=20	Paired Samples t-test	2.0 (SD=1.3)	3.5 (SD=1.2)	1.5	< 0.001	t(19) = -4.65	0.53
Pre-Post VR Basic Understanding N=11	Paired Samples t-test	2.2 (SD=1.3)	4.2 (SD=1.2)	2.0	0.001	t(10) = -4.69	0.69
TT Post Survey One Play Basic Understanding N=45	Paired Samples t-test	1.9 (SD=1.6)	3.1 (SD=1.5)	1.2	< 0.001	t(44) = -7.06	0.53
VR/Tablet Post Survey One Play Basic Understanding N=26	Paired Samples t-test	1.8 (SD=1.3)	4.1 (SD=0.9)	2.3	< 0.001	t(25) = -8.83	0.73

Data Analysis. Data analysis employed social science research methods, drawing from approaches used by other researchers to investigate similar informal science learning experiences (summarized in Cors, 2016). The analysis followed an iterative process. Each cycle started with a basic analysis of survey data, conducted by investigator Cors, according to the methods described below, which was followed by an interpretive session with other team members. Most of these interpretive sessions occurred with core team members. Two interpretive exchanges took place with the larger IceCube team. On May 15th, 2018, for example, the IceCube team looked at early survey results, to decide whether to change data collection strategies. In November 2018, the IceCube team was asked via email to comment on results from analysis of the main sample of 154 subjects. Subsequent data analysis decisions were shaped by these discussions.

Analysis of Subject Ratings. To analyze responses to Likert scale survey data, the team used the statistical software package IBM SPSS. Researchers used paired sample t-tests to compare subjects' responses about their pre-game and post-game experiences. Independent t-tests were used to compare responses from subjects who tried the TT experience with those who tried the VR experience. To compare three different subject groups, based on which games they played, a one-way ANOVA was used.

Analysis of Concept Map and Multiple-Choice Responses. Performance on concept map and multiple-choice survey items was calculated as a correctness score, which is the percentage of subjects who answered correctly (# correctly answered/ # answered). For the multiple-choice items, a score was considered a correct score when a subject chose the right answer among four answers. For the concept maps, a score was considered correct when a subject moved (selected) four of five correct concepts into the map.

HYPOTHESES

Using the two experiences and developed research methods surrounding the exhibit, the team wanted to understand the effectiveness of VR and TT technology for promoting public understanding about IceCube. Specifically, researchers wanted to understand how subjects experienced the exhibit, including their knowledge and curiosity, what they took away from the exhibit, and what factors affected their experience. These research aims were based on broad intended outcomes for the exhibit to improve public understanding of IceCube and also on more focused intended visitor outcomes of basic understanding of, and curiosity about, IceCube. Furthermore, given that VR was seen as being more novel than the TT, it was thought that the VR experience would generate greater interest and engagement compared to the TT experience. Considering these goals, and also previous research, the team developed the following hypotheses regarding the exhibit:

- A. People who use the exhibit will gain a basic understanding of the IceCube project.
- B. There will be a difference in the basic understanding gained between those that used the TT, and those who used VR.
- C. Use of the VR experience will spark the curiosity of people in IceCube more strongly than use of the TT.
- D. People who perform better (achieve a higher score) at the exhibit games will gain a better understanding of IceCube than those whose game scores are lower.

RESULTS

Hypothesis A: Gain in Basic Understanding. The research team analyzed responses to three of the survey questions to understand whether users achieved a gain in basic understanding of IceCube and neutrinos. The first analysis of data to test Hypothesis A examined responses from the main post sample (N=154). The team employed a paired samples t-test to compare two ratings provided by the subjects after playing a game: their basic understanding of IceCube on a scale of 0 (None) to 5 (Expert) prior to trying each experience (a retrospective rating), and then again after trying each experience. Findings support Hypothesis A, as perceived basic understanding increased significantly from before to after trying each experience (Table 2). That is, gains in perceived basic understanding rose significantly for subjects who completed the TT survey (“TT” means they likely played the TT exhibit) from before to after playing the game with a mean gain of 1.4 ($p < 0.001$). For subjects who completed the survey on the iPad next to the VR headset (“VR” means they likely played the virtual reality game), mean basic understanding also rose significantly with a mean gain of 2.1 ($p = 0.002$). This suggests a perception or feeling of having better understood IceCube and the eta squared statistics indicate there was a large effect size for both experiences, based on the guidelines proposed by Cohen (Cohen, 1992): .01=small effect, .06=moderate effect, .14=large effect. To note: in all post-sample survey analyses, subjects may have played one or both experiences. TT and VR tablet refer to how the surveys were filled out, either integrated in the TT or on the tablet next to the exhibit.

Significant gains in basic understanding of IceCube were also evident in results from a paired samples t-test of pre-post sample data (actual before and after responses, N=31), with large effect sizes (Table 2). Specific results show that for subjects who played the TT (N=20), perceived understanding rose by a gain of 1.5 ($p < 0.001$). For those experiencing the VR (N=11), perceived basic understanding rose

Table 3. Difficulty scores from Pre-Post sample (N=31) on performance tests show, with one exception, (actual) increases in game players’ basic understanding.

Pre-Post Sample (N=31)	Correctness Score Before Game	Correctness Score After Game
Ice Cube Multiple Choice	84%	84%
Neutrino Multiple Choice	65%	81%
IceCube Concept Map	39%	42%
Neutrino Concept Map	65%	74%

by a mean gain of 2.0 ($p = 0.001$). The pre-post sample survey results hold that persons who filled out the TT survey, only played the TT, while persons who filled out the tablet survey, only played the VR experience.

The team was also interested in comparing outcomes where people played either experience only once, to better understand what someone gains from a single play. To analyze results from this ‘one-play’ subset of subjects, the research team segmented out post-sample results where a person indicated on the survey that they only played the TT once, or the VR once, without interacting with the other system. The one-play sample consisted of 71 total people. Of those 71 people, 45 people filled out the TT survey once, and 26 people filled out the VR survey once. In the sample, 46 people reported female, 24 people reported male, and 1 reported other. About half of the subjects were between ages 18-30. Results from a paired samples t-test support Hypothesis A, showing significant gains in perceived basic understanding with large effect. Mean ratings in basic understanding from subjects who filled out the TT rose by 1.2 ($p < 0.001$). Ratings from those who filled out the VR rose by 2.3 ($p < 0.001$) (Table 2).

Performance measures for multiple choice and concept map items from both pre and post surveys (both N=31) were calculated using aforementioned correctness score, which is

Table 4. Independent t-test results show that players of VR gave higher average ratings for basic understanding than players of TT. Curiosity differed between the two, with the VR showing a slightly higher mean, but not significantly higher than the TT (for both the overall, and one play cases).

Factor	Test	VR mean N=75	TT mean N=79	p	t	η ²
Change in perceived basic understanding (retrospective rating)	Independent t-test	M=2.1 (SD=1.4)	M=1.3 (SD=1.3)	0.001	(t(152) =3.45)	0.072
Curiosity	Independent t-test	M=4.7 (SD=1.4)	M=4.4 (SD=1.7)	0.23	--	--
		VR mean (one play) N=26	TT mean (one play) N=45			
Curiosity	Independent t-test	M=5.0 (SD=1.1)	M=4.5 (SD=1.6)	0.10	--	--

the percentage of subjects who answered correctly (# correct answers / # answered). These findings, like the perceived basic understanding data, showed increases in basic understanding, with one exception. The only item for which subject performance showed no gain was on the multiple-choice question, “Where is Ice Cube?”, where 26 of 31 subjects answered correctly (26/31 = 84% correctness) both before and after playing the exhibit. Such a high score that did not change from before to after playing the exhibit suggests the question may have been too easy, creating a ceiling effect. The remaining test results showed increases in basic understanding scores from pre to post. That is, correct responses to the multiple-choice question about neutrinos increased from 20 (20/31=65%) to 25 (25/31=81%). Subject performance also improved on the concept map questions. Correct responses to the IceCube concept map increased from 12 (12/31=39%) to 13 (13/31= 42%) and correct responses on the neutrino concept map increased from 20 (20/31=65%) to 23 (23/31=74%). Pre-post survey results are summarized in Table 3.

Hypothesis B: Difference in Basic Understanding between Exhibits. The research team wanted to compare the two experiences in terms of how they promoted subjects’ gains in basic understanding of IceCube. The research team conducted several analyses that support Hypothesis B. An initial analysis of main sample data (N=154) employed an independent samples t-test to compare the means for gains in basic understanding from players of the VR with players of the TT. Results show players of the VR responded with significantly greater perceived gains in basic understanding of IceCube. Specific test results show VR players (N=75) had mean gains of M=2.1, which is greater than TT players (N=79) mean gains of M=1.3. According to the statistic, the average gain in basic understanding from the VR players was significantly greater than the mean gain reported by TT players ($t(152)=3.45$, $p=.001$, 2-tailed), with moderate effect (eta squared=.072).

Similarly, a look at one-play sample (N=71) data, irrespective of which survey they filled out, showed that VR players who did not play the TT had significantly greater gains in perceived basic understanding of IceCube than one-time TT players who did not play the VR. These findings were derived from a one-way ANOVA to compare gains in basic understanding among subjects in three groups: (1) subjects who played the VR once (and not the TT) (N=24); (2) subjects who played the TT only once (and not the VR) (N=47), and (3) subjects who played both the VR and the TT at least once (N=34). Test results showed subjects who played the VR once had the greatest mean gain in perceived basic understanding: (1) M=2.4 (SD=1.3), (2), M=1.2 (SD=1.6), (3) M=2.1 (SD=1.5). There was a statistically significant difference at the $p<.05$ level of the ratings

($F(3,150)=5.4$, $p=0.001$), which had moderate-high effect size (eta squared=0.10). Post Hoc test results show that responses from players of the VR only group (1), varied significantly from results of players of the TT only group (2), ($p=0.003$). Also, responses from players of the TT only, group (2) varied significantly from those from subjects who played both games at least once, group (3), ($p=0.022$).

Hypothesis C: Curiosity Sparked Would Be Greater for VR than for the TT. Post-sample ratings from subjects show that regardless of which experience a user played, they were curious to learn more about IceCube after playing. Curiosity ratings on a scale of 1 (not at all curious) to 6 (extremely curious) showed an average of M=4.7 from those who played the VR and M=4.4 from those who played the TT. The difference, according to results of an independent samples t-test, was not significant ($p=0.23$) (Table 4). Similar results were found from an analysis with the one-play sample (N=71), where subject ratings suggest that the curiosity of subjects who played the VR once was sparked to a greater degree (M=5.0) than for those who played the TT once (M=4.5), although not significantly greater ($p=0.10$) (Table 4). Interview data during field tests of the exhibit also show that subjects’ curiosity is sparked by either exhibit. That is, specific suggestions in open-ended comment answers included suggestions such as adding extra physical reading material (such as pamphlets, brochures, or posters) about the subject matter, offering additional evidence that subjects feel curious and want to know more about IceCube.

Hypothesis D: Higher Scores Would Result in Greater Gains in Basic Understanding. While the VR was an interactive experience, the concepts of scores was not implemented meaning that Hypothesis D could only be tested on the TT experience. In means of testing hypothesis D, the team gathered scores throughout the playing process, and matched timestamps from these scores to timestamps on survey results. Findings from the post-sample support the null hypothesis, which states that the scores on the TT will not correlate significantly with basic understanding. The only correlation that was found to be moderately significant was between number of touches while playing the TT and perceived basic understanding before the game (retrospectively rated) ($r=.3$, N=76, $p=.002$). This result could indicate that those subjects who thought they knew a lot about IceCube before playing the game(s) were encouraged by their perceived knowledge to more boldly try different game options, which was detected by a higher frequency in screen touches.

DISCUSSION

Research Findings. Overall, the results demonstrated that users of the exhibit came away from the experiences with

a greater knowledge of neutrinos and the IceCube project, confirming Hypothesis A. The findings related to Hypothesis A contribute to a methodological debate (summarized in Hill and Betz, 2005) about the usefulness of retrospective posttest (RPT) survey item designs, which ask subjects to provide ratings after an intervention about their opinion or feeling before and after that intervention. Our findings show that RPT results showing perceived gains in basic understanding were validated by traditional (actual) pre- and posttest (TPT) perceived basic understanding data, and also validated by actual gains in performance on multiple choice and concept map tests. These findings demonstrate that RPT designs offer rigor in data collection that is as good as TPT when studying visitor experiences with interactive exhibits, something impactful for resource-poor museums. Using TPT research designs to collect survey data about user outcomes from an interactive exhibit is often cost-prohibitive for informal learning places such as museums and it also introduces challenges in obtaining complete responses from visitors. Using a RPT, where exhibit users give their before-intervention rating only once, and again after interacting with the exhibit, addresses these concerns. Previous studies also show that the RPT method reduces response-shift bias and recall bias (Hill and Betz, 2005), something especially useful with short interventions, such as interactive exhibits.

When directly comparing the learning outcomes between the two experiences, Hypothesis B, showed that the VR experience resulted in a significantly greater gain in basic understanding. This could be for many reasons: the VR experience was longer than the TT (3 minutes vs. 60-75 seconds), and also incorporated voice-over audio and captions, therefore allowing for a different, more experiential form of learning. In addition, a user potentially became more engaged in the experience through VR's ability to create a sense of presence (Slater, 1999). Hypothesis C, which predicted there would be greater curiosity sparked from the VR than the TT, did not hold true, as there was not a significant difference in the amount of curiosity sparked between the two experiences. An additional hypothesis that did not prove significant was hypothesis D regarding higher scores leading to higher learning outcomes. While on the one hand this result suggests a decoupling of gameplay elements and learning, this may not be entirely negative as it can be surmised that individuals who were not skillful in the gameplay setting were still able to show a knowledge gain. It may be that learning was accomplished quickly within the gameplay setting and therefore increasing one's point total was not useful in increasing one's knowledge of the content. Future work will aim to better understand at which moments during gameplay learning occurs.

Limitations. While the collected data shows generally positive results, there are several points worth further discussion.

One aforementioned evaluation type that was administered in the surveys is known as the concept map (Markham et al., 1994). One aim of the pilot study was to test and develop two concept maps for use in measuring game players' basic understanding of IceCube. Concept maps are a kind of knowledge tests where experts should consistently score perfectly almost every time, while learners should ideally score below 50% before an intervention / program and, if the program is effective, above 50% after. If learners score on average greater than about 80% on a pre-test, this suggests there is little room left for learning, or a ceiling effect.

Concept map scores were assessed according to how difficult they were for subjects. This measure is called 'Correctness' and measures the percentage of subjects who answered correctly. According to Correctness scores from pre-post sample data, shown in Table 3, all of the questions became less difficult except for the questions, "Where is IceCube?" The high pregame score of 84%, and the absence of change in the score, suggests a ceiling effect and indicates that the question is too easy.

One subject that provided some controversy was the term "radiation" in one of the concept maps. While the term was meant to be unrelated to the IceCube detector, which indirectly detects neutrinos as opposed to radiation, there is a relationship between neutrinos and radiation elements. Despite the term radiation not appearing in either exhibit game, many respondents included it in their concept map. Future studies will use a different, more discerning term to avoid this confusion.

A major limitation of the study was the lack of data generated from individuals who experienced the exhibit, but did not take the survey. While raw metrics, such as play counts, were gathered from these experiences, learning outcomes were not assessed. Future work will aim to develop passive methods to measure learning outcomes and attributes of the exhibit such as drawing and holding power (Serrell, 2002). As an additional limitation, since the exhibit was only deployed in a single space on a university campus, the diversity of participants was likely not as high as if the system was deployed in multiple settings. Future work aims to resolve this by creating deployable and downloadable experiences that can reach a broader audience.

Another limitation of the study occurred due to a selection bias of subjects. This stems from the fact that the exhibits were installed in an informal environment within a building on a University campus. Since the exhibits are located on a University campus, many subjects fell into the 18-25-year-old range. The research team recognized this limitation in the early stages of deployment, but kept the exhibit at its location for a variety of reasons. The first reason is that a partnership with another entity in the same building required the exhibit to stay at its location. A second reason was that once a month, a building event occurs that encour-

ages families and members of the public to come explore a science-related theme through several informal outreach activities. These events gather a large, diverse audience, although typically young.

Recruitment for the pre-post sample occurred via directly asking passers-by if they were interested in participating in a study. The team worked together with Discovery Outreach to conduct the pre-post recruitment on a day when there was known to be a greater number of people on the same floor of the building (a research poster session). Persons who administered the pre- and posttest felt that conducting the recruitment on a day that coincided with this additional event was very helpful in attracting passers-by to participate in the study and try out the exhibit. Without coinciding the pre-post testing on a day of additional events, recruitment in an informal setting becomes much more challenging.

An additional limitation of the study relates to the fact that the post-sample lacks a true baseline measurement. Part of this limitation is due to the general challenge of recruiting subjects in an informal context. The team developed the exhibits to run 24 hours a day, seven days a week, in the publicly accessible space of the WID. For the TT, a button pops up on completion of the game, giving the player an option to complete a survey. While the team could have also had this button pop-up right away prior to playing the game in efforts to obtain baseline data, chances are persons may have skipped over this with a high frequency (only roughly 3% of persons even filled out the post-survey). For the VR, a stand-alone pre-survey becomes a greater logistical challenge, due to the fact that a person needs to wear the HMD to experience the content, blocking the user's view of the real world. As a future approach, two possible solutions exist that could direct the user to fill out a pre-survey for the VR. One, a message could be programmed into the experience to opt-in to participating in the study, and if so, direct the user to fill out the pre-survey on an adjacent iPad. Upon filling out the survey, a user could return to the HMD, experience the VR, and then return to the iPad to fill out a post survey. The HMD experience would require slightly different programming than what the research team implemented, as any time a person takes off the HMD, the VR scenario resets to the beginning (to facilitate the exhibit to operate independent of docents 24 hours a day, seven days a week). The other option would be to fully implement the surveys within the VR via a 3D user interface to present questions and allow the user to choose responses directly within the VR environment.

A limitation in the design of the evaluation instrument stems from the fact that the TT exhibit shows a welcome message at the top of the screen that mentions the South Pole, while one of the questions in the evaluation instrument asks where IceCube is located. Therefore, people may have been able to answer this question, potentially without ever experiencing the exhibit. Future work would avoid using a

question that can be answered without experiencing the exhibit. The limitation of gleaning information from the exhibit without actually playing it is elaborated on below.

One can argue that there exists a general challenge and potential evaluative limitation that can occur when users see a different person's game play experience prior to playing themselves. This would occasionally occur while waiting and watching the game on a monitor mounted within the exhibit's physical structure (Figure 5), or if people look over the shoulder of someone playing the TT. The question is raised: how much did this preview of the game affect or improve their basic understanding? While this may have contaminated subjects' basic understanding responses on the study surveys, investigators noticed that this additional way of experiencing the exhibit has opportunities of its own. Specifically, it can promote discussion among observers of IceCube and can reinforce basic understanding of the project, promoting public awareness, which is a key desired outcome for the exhibit.

Orientation: Challenges and Solutions. The development team faced several challenges during the design and testing stages when deploying the exhibit in the public research space. Each portion of the exhibit needed to run independently without any guidance. In this vein, orientation of users to the exhibit was crucial, and the team made specific design decisions to accommodate user orientation (Griffin, 1994). To help with orientation in the TT, the game shows a tutorial when no user is interacting with the TT. The tutorial loops a sequence of instructional and reference information pertaining to IceCube and neutrinos. In addition to providing basic instructions, a tutorial mode allows passers-by who may not be interested in playing, the opportunity to glean information about the subject-matter. Upon pressing the TT, the game provides an optional choice to see a sequence of instructions. If the user chooses to see instructions, nine instructional panels provide basic information on how to play the game. These instructions guide the user through an example swipe action to detect a neutrino. This sequence teaches the player what they need to do by practice before it counts towards their official score. In this way the player is acquainted with the gameplay interactions before adding time pressure. Following this trial, a sixty second timer starts and game play begins.

To aid in orientation for the VR exhibit, the user chooses a language (English, Spanish or Portuguese) to start the narration to orient the user. During the first few moments, the narrator tells the user to look around, and gaze at specific targets that are up and down from their typical forward point of view. This prompting teaches the user that they can look around in the HMD to experience the Antarctic environment all around them. After the orientation, during later points of the experience, directions are given via captions

and narration explaining to the user how to advance through the experience.

Future Directions. The research team believes in several future directions for the presented work. While the research shows that the exhibit successfully gives people a basic understanding of neutrinos and IceCube, ideally, future work would integrate the learning assessment directly into the experience as opposed to a survey. In an even more ideal situation, the VR and TT experiences would allow users to obtain a knowledge of neutrinos and IceCube, by performing actions within the experiences that don't explicitly provide as much information on the subject-matter through text. Rather, a user may learn naturally through intended interaction with the exhibit, allowing the user to form their own hypothesis and explore them repeatedly, i.e. the probing principle (Gee, 2003).

Another future direction would be a more direct comparison between the technologies. While the original intention of the project was to build a mutual experience between the TT and VR environments, it was determined early in the development cycle that the affordances of the two systems were too different. This, in turn, led to the development of experiences tailored to the features of each system. While this decision created two experiences both able to produce effective outcomes, the impact of each aspect that makes up the overall experience remains unclear. Future work aims to further understand what components make up an effective experience.

Finally, the team would like to study the attracting and holding power of the exhibit. It would be useful to know what drew people to the exhibit and what barriers prevented others from interacting with the exhibits. The ultimate goal is to determine how to reach and engage new audiences that are normally passed over.

CONCLUSION

The developed experiences successfully created a new informal STEM learning exhibit of the highly complex astrophysical topics of neutrinos and their detection at the IceCube observatory in Antarctica. This research study showed that both exhibits increased basic understanding of neutrinos and IceCube. While the VR participants showed a greater increase in basic understanding compared to the tabletop exhibit, both exhibits demonstrated the ability to spark the curiosity of participants. Both experiences demonstrated the ability to spark the curiosity of participants. Future work aims to better understand what effects the different technologies have to offer, to test amongst a more diverse sample, and to create evaluation mechanics built into the experience, thus enabling passive evaluation of learning outcomes.

ASSOCIATED CONTENT

Supplemental material mentioned in this manuscript can be found uploaded to the same webpage as this the manuscript.

AUTHOR INFORMATION

Corresponding Author

Ross Tredinnick. rdtredinnick@wisc.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

FUNDING SOURCES

This project has been supported through NSF Award #1612504, the Wisconsin Institute for Discovery, the Wisconsin Alumni Research Foundation, and WIPAC.

ABBREVIATIONS

APE: Active Prolonged Engagement; DOM: Digital Optical Modules; HMD: Head Mounted Display; NSF: National Science Foundation; TT: Touch Table; VR: Virtual Reality; WID: Wisconsin Institute for Discovery; WIPAC: Wisconsin IceCube Particle Astrophysics Center.

REFERENCES

- Block, F., Horn, M. S., Phillips, B. C., Diamond, J., Evans, E. M., and Shen, C. (2012). The deeptree exhibit: Visualizing the tree of life to facilitate informal learning. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2789-2798.
- Ciocca, G., Olivo, P., and Schettini, R. (2012). Browsing museum image collections on a multi-touch table. *Information systems*, 37(2), 169-182.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Correia, N., Mota, T., Nóbrega, R., Silva, L., and Almeida, A. (2010, November). A multi-touch tabletop for robust multimedia interaction in museums. In *ACM International Conference on Interactive Tabletops and Surfaces* (pp. 117-120). ACM.
- Cors, R. (2016). *Informal Science Learning: an investigation of how novelty and motivation affect interest development at a mobile laboratory* (Doctoral dissertation, University of Geneva).
- Falk, J. H., Martin, W. W., and Balling, J. D. (1978). The novel field-trip phenomenon: Adjustment to novel settings interferes with task learning. *Journal of Research in Science Teaching*, 15(2), 127-134.

- Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1), 20-20.
- Griffin, J. (1994). Learning to learn in informal science settings. *Research in science education*, 24(1), 121-128.
- Grafe, M., Wortmann, R., and Westphal, H. (2002). AR-based interactive exploration of a museum exhibit. In *The First IEEE International Workshop Augmented Reality Toolkit*, (pp. 5-pp). IEEE.
- Hill, L. G., and Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501-517.
- Hirose, M. (2015). Virtual reality technology and museum exhibit. *International Journal of Virtual Reality (IJVR)*, 5(2), 31-36.
- Horn, M., Atrash Leong, Z., Block, F., Diamond, J., Evans, E. M., Phillips, B., and Shen, C. (2012, May). Of BATs and APES: an interactive tabletop game for natural history museums. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2059-2068). ACM.
- Hornecker, E., and Stifter, M. (2006, November). Learning from interactive museum installations about interaction design for public settings. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (pp. 135-142). ACM.
- Hornecker, E. (2008, October). "I don't understand it either, but it is cool"-visitor interactions with a multi-touch table in a museum. In *2008 3rd IEEE international workshop on horizontal interactive human computer systems* (pp. 113-120). IEEE.
- Humphrey, T., and Gutwill, J. P. (2017). *Fostering active prolonged engagement: The art of creating APE exhibits*. Routledge.
- IceCube Collaboration. (2013). Evidence for high-energy extraterrestrial neutrinos at the IceCube detector. *Science*, 342(6161), 1242856.
- IceCube, T. (2018). Multi-messenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A. arXiv preprint arXiv:1807.08816.
- Lepouras, G., and Vassilakis, C. (2004). Virtual museums for all: employing game technology for edutainment. *Virtual reality*, 8(2), 96-106.
- Markham, K. M., Mintzes, J. J., and Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. *Journal of research in science teaching*, 31(1), 91-101.
- Mohapatra, R. N., and Pal, P. B. (2004). *Massive neutrinos in physics and astrophysics* (Vol. 72). World scientific.
- Podgorny, J. (2004). *Studying visitor engagement in Virtual Reality based children's science museum exhibits*. Unpublished master dissertation, Department of Astronomy and Astrophysics, University Of Chicago, Chicago, IL.
- Qualtrics, L. L. C. (2014). *Qualtrics [software]*. Utah, USA: Qualtrics.
- Roussou, M. (2001, September). The interplay between form, story, and history: The use of narrative in cultural and educational virtual reality. In *International Conference on Virtual Storytelling* (pp. 181-190). Springer, Berlin, Heidelberg.
- Serrell, B. (2002). Are they watching? Visitors and videos in exhibitions. *Curator: The Museum Journal*, 45(1), 50-64.
- Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence*, 8(5), 560-565.
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109(3), 431-433.
- Stogner, M. B. (2011). "The Immersive Cultural Museum Experience--Creating Context and Story with New Media Technology". *International Journal of the Inclusive Museum*, 3(3).
- Tost, L. P., and Economou, M. (2007). Exploring the suitability of Virtual Reality interactivity for exhibitions through an integrated evaluation: the case of the Ename Museum. *Museology*, 4, 81-97.
- Van Dijk, E. M., Lingnau, A., and Kockelkorn, H. (2012, October). Measuring enjoyment of an interactive museum experience. In *Proceedings of the 14th ACM international conference on Multimodal interaction* (pp. 249-256). ACM.
- Wojciechowski, R., Walczak, K., White, M., and Cellary, W. (2004, April). Building virtual and augmented reality museum exhibitions. In *Proceedings of the ninth international conference on 3D Web technology* (pp. 135-144). ACM.
- York, D. G., Adelman, J., Anderson Jr, J. E., Anderson, S. F., Annis, J., Bahcall, N. A., ... and Boroski, W. N. (2000). The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3), 1579.
- Zaharias, P., Michael, D., and Chrysanthou, Y. (2013). Learning through multi-touch interfaces in museum exhibits: An empirical investigation. *Journal of Educational Technology and Society*, 16(3), 374-384.