MEASURING THE EFFECT OF USER EXPERIENCE AND ENGAGEMENT ON LEARNING USING INTERACTIVE SIMULATIONS

A Dissertation
Presented to
The Academic Faculty

by

Brianna J. Tomlinson

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Human-Centered Computing in the
School of Interactive Computing

Georgia Institute of Technology
August 2020

MEASURING THE EFFECT OF USER EXPERIENCE AND ENGAGEMENT ON LEARNING USING INTERACTIVE SIMULATIONS

Approved by:

Dr. Bruce N. Walker, Advisor
School of Interactive Computing
Georgia Institute of Technology

Dr. Marion Usselman
Center for Education Integrating Science, Mathematics and Computing (CEISMC)
Georgia Institute of Technology

Dr. Betsy DiSalvo
School of Interactive Computing
Georgia Institute of Technology

Dr. Emily B. Moore
Department of Physics
University of Colorado Boulder

Dr. Maribeth Coleman
School of Interactive Computing
Georgia Institute of Technology

Date Approved: July 16, 2020

To my mom and dad, Patricia and Peter Tomlinson, who encouraged me to pursue whatever I wanted and who taught me to always be curious and love learning. To my sisters, Alyssa and Celinna, and Joel, Rachel, and Brittany for all of the cat pictures. To Cheryl and Jereme for their support, even when life takes us far away. To Tia for her never-ending encouragement. To all of my friends and family, you mean so much to me. To my partner, Joshua Lapinski, for helping me learn when to take a break, and for all of our cooking and travel adventures.

ACKNOWLEDGEMENTS

The list of people who have supported me over the course of my time at Georgia Tech, and over the duration of my dissertation is extensive, and I will do my best to highlight everyone.

First, to April Whitt, from the Fernbank Science Center – this project would have been impossible without your support and encouragement early on. To CEISMC, for helping me recruit instructors for the initial interviews and for supporting my participant recruitment efforts. To Giant Army, for providing access to the Universe Sandbox code; without the support from Dan Dixon and Jacob Williams, my coding efforts would have seemed insurmountable. To Mike Winter, Chris Latina, Smruthi Bhat, and Milap Rane for the help in creating the solar system sonification.

Next, to PhET Interactive Sims: without the 3+ years of experience working with your team, I may never have considered studying interactive simulations for my dissertation. It was amazing to work with you all, and I learned so much from our projects.

To my committee members, who took time to meet, mentor, and guide my dissertation experience. Betsy, Emily, Maribeth, and Marion: I appreciate your advice, and also your flexibility when unforeseen circumstances changed data collection possibilities in Spring 2020. To my advisor, Bruce Walker, for letting me explore a variety of projects, encouraging me to do more and also when to do less, and for helping me become the researcher I am today.

I would like to thank the faculty and students from Human-Centered Computing and GradWomen@CC, and all of the friends I've made along the way for their support (including Ari, Synge, Udaya).

To Tech Taido, for giving me a place to relax, not think about work, and learn the joy of Taido. In particular, I would like to thank Bryan-sensei, for starting Tech Taido. Grad school would have been so lonely without my Taido family. Being a part of Tech Taido has been one of the great joys of graduate school! I cannot begin to explain how important it has been to me and how large of an impact you have had on my life; particularly, I would like to thank Bryan, Laura, Bolot, Lacy, Rahul, Teresa, Thomas, Manan, Jason, Tomer, Leo, Chatfield, Megan, Alex, Emily, Gwyn, Logan, Malik, Varun, Ressa, Nicole, Vineeth, Josh, Grace, Lauren, Matthew, Thomas, Sarah, Nolan, DeAndre, Haniel, Theresa, Shelley, Ronald, and everyone from MSU and the Honbu.

To the numerous friends I made working in the Sonification Lab. I could not have asked (or hoped) for better lab mates. I've learned so much from you and working with you has helped me become a better mentor, researcher, and person. To the rants, the research rabbit holes, the serious conversations, the rejections, the acceptances, the listening, the learning, and most of all to your support and friendship; I would like to thank (in no particular order, mind you!): Jonathan, Jared, Thom, Carrie, Vincent, Brandon, Stanley, Heather, Nadia, Zoe, Brittany, Rachel, and Keenan.

I would like to thank my parents, sisters, friends, and family for their support over the past seven years. I could not have done this without your endless encouragement and love. Thank you for your patience when school took precedence and when my time was

short. And finally, to Josh, I realize that the late nights and long hours were difficult and draining, but your composure, support, and programming advice mean more than you know.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS **iv**

LIST OF TABLES **xii**

LIST OF FIGURES **xiv**

LIST OF SYMBOLS AND ABBREVIATIONS **xv**

SUMMARY **xvi**

CHAPTER 1. Introduction and Motivation **1**
1.1 Motivation **1**
1.2 Thesis Statement **2**
1.3 Research Questions **2**
1.4 Summary of Studies **3**

CHAPTER 2. Background **5**
2.1 Learning Theory **5**
2.1.1 Learning 5
2.1.2 Understanding and Learning in Context 7
2.1.3 Multimedia Learning 13
2.2 Educational Technology **15**
2.2.1 Interactive Simulations 16
2.2.2 Virtual Reality (VR) 17
2.2.3 Auditory Displays 21
2.3 Educational Technology Evaluations **23**
2.3.1 Learning Studies in Interactive Simulations 23
2.3.2 Learning Studies in Virtual Reality 25
2.3.3 Human-Computer Interaction (HCI) Evaluations 30
2.4 Educational Topic Focus **32**
2.4.1 Next Generation Science Standards (NGSS) 32
2.4.2 Common Core 33
2.4.3 Space Simulations 34

CHAPTER 3. Preliminary Work **35**
3.1 Study 1: Planetarium Study **35**
3.1.1 Teacher Interviews 36
3.1.2 Misconception Identifier Study 38
3.1.3 Solar System Sonification Design 42
3.2 Study 2: Listening Activity Evaluation **50**
3.2.1 Purpose 50
3.2.2 Methods 51
3.2.3 Analyses and Results 53
3.2.4 Discussion 54

3.2.5 Conclusion 55

CHAPTER 4. Study 3 **58**
4.1 Participants **58**
4.2 Materials **59**
4.2.1 Universe Sandbox Designs 59
4.2.2 Pre-Activity Surveys 65
4.2.3 Activity 66
4.2.4 Post-Activity Surveys 66
4.2.5 Follow-up Interview 67
4.3 Procedure **67**
4.4 Pilot Testing **69**
4.5 Research Design **71**
4.5.1 Hypotheses 72
4.5.2 Analyses 73

CHAPTER 5. Study 3 Results **76**
5.1 Research Question 1 **78**
5.1.1 Physical Engagement 79
5.1.2 Emotional Engagement 82
5.1.3 Cognitive Engagement 83
5.1.4 Motivation 88
5.1.5 Long-term Engagement 90
5.1.6 Qualitative Observational Results 92
5.2 Research Question 2 **100**
5.2.1 Cognitive Engagement 101
5.2.2 Long-term Engagement 103
5.2.3 Usability Scores 104
5.2.4 Emotional Engagement 105
5.2.5 Motivation 106
5.2.6 Physical Engagement 107
5.2.7 Qualitative Observational Results 109
5.3 Research Question 3 **113**
5.3.1 Technology Use 113
5.3.2 Forecasted Affect 115
5.3.3 Additional Science Activities 116
5.3.4 Learner Perception Towards Science 117

CHAPTER 6. Study 3 Discussion **119**
6.1 VR and PC Differences **119**
6.2 Audio and No-audio Differences **123**
6.3 Practical Implications for Educational Simulations **127**
6.4 Application Domains **132**
6.5 Limitations and Participant Considerations **134**
6.6 Future Research **135**
6.7 Conclusion **136**

CHAPTER 7. Discussion **138**
7.1 Interaction's Effect on Sonification Interpretation **138**
7.2 Sound Literacy **141**
7.3 When is VR or Audio Particularly Relevant? **143**
7.4 Multimedia Principle **146**
7.5 Evaluating Auditory User Experience **147**
7.6 What is Engagement, Really? **149**
7.7 Conclusion **151**

APPENDIX A. Science Teacher Interviews **153**
A.1 Semi-structured Question Guide **153**
A.2 Demographics Questions **153**

APPENDIX B. Misconception Identifier Survey **156**
B.1 Misconception Survey Questions **156**
B.2 Exit Survey Questions **157**

APPENDIX C. Solar System Sonification **159**
C.1 Recording **159**
C.2 Script Outline **159**
C.3 Planetarium Survey Questions **161**

APPENDIX D. Listening Activity Materials **164**
D.1 Pre/Post Test **164**
D.2 Listening Activity Questionnaire **169**
D.3 Demographics **172**
D.4 BUZZ: Audio User Experience Questionnaire **172**

APPENDIX E. Pre-Lab Activity Materials **174**
E.1 Adapted Technology Experience Profile **174**
E.2 Modified Attitudes Toward Science Instrument (mATSI) **177**
E.3 Positive and Negative Affect Schedule (PANAS) **181**
E.3.1 Group A Prompt 181
E.3.2 Group B Prompt 181
E.3.3 Group C Prompt 181
E.3.4 Group D Prompt 181
E.3.5 PANAS questions 182
E.4 Pre-Activity Solar System Questions **185**
E.5 Sim Sickness Questionnaire (SSQ) **186**

APPENDIX F. Lab Activity Materials **188**
F.1 Prompt Set 1 & 2 **188**
F.2 Prompt Set 3 **188**
F.3 Prompt Set 4 **188**

APPENDIX G. Post Activity Materials **189**
G.1 Post-Activity Solar System Questions **189**

G.2 Positive and Negative Affect Schedule (PANAS) 190
G.2.1 Group A Prompt 190
G.2.2 Group B Prompt 190
G.2.3 Group C Prompt 190
G.2.4 Group D Prompt 190
G.2.5 PANAS questions 190
G.3 Slater-Usoh-Steed Presence Scale (SUS) 191
G.4 Usability Metric for User Experience (UMUX) 194
G.5 Audio User Experience Scale (BUZZ) 195
G.6 Motivated Strategies for Learning Questionnaire (MSLQ) 202
G.7 Science Activity Questionnaire (SAQ) 206
G.8 Open-ended Questions 211
G.9 Demographics 212

APPENDIX H. Follow-Up Interview Materials 214
H.1 Interview Questions 214

APPENDIX I. Adapted Survey Materials 215
I.1 SAQ Original and Adapted 215
I.2 MSLQ Original and Adapted 215

APPENDIX I. Process Code Definitions 218
J.1 Interacting with Planet Information and Details 218
J.2 Using Movement or View Controls 218
J.3 Verbalizing Comments and Observations 219
J.4 Playing or Replaying Sounds 220
J.5 Completing Non-Task Play 220

APPENDIX K. Process Coding Counts And Statistics 221
K.1 Code Counts 221
K.1.1 Group A Codes: PC – no-audio 221
K.1.2 Group B Codes: PC – audio 223
K.1.3 Group C Codes: VR – no-audio 225
K.1.4 Group D Codes: VR – audio 227
K.2 Chi-square Observed 229
K.2.1 Top-level Code Categories 229
K.2.2 Viewing Details Only 230
K.2.3 PC-Only Comparisons 231
K.2.4 VR-Only Comparisons 232
K.2.5 Verbalizations 233

APPENDIX L. Post-Activity Free Response 234
L.1 Question 1 234
L.2 Question 2 236
L.3 Question 3 237
L.4 Question 4 238
L.5 Question 5 239

L.6 Question 6 **240**

APPENDIX M. Follow-up Interview Responses **241**
M.1 Question 1 **241**
M.2 Question 2 **243**
M.3 Question 3 **244**
M.4 Question 4 **245**
M.5 Question 5 **246**
M.6 Question 6 **247**
M.7 Question 7 **248**
M.8 Question 8 **249**
M.9 Question 9 **250**
M.10 Question 10 **251**
M.11 Question 11 **252**

REFERENCES **253**

LIST OF TABLES

Table 1	- Reliability summary table for the overall score and the SSP and PP individual factors.	54
Table 2	- The four study conditions.	72
Table 3	- Pre-Activity Average Scores in order of participant completion: technology use (frequency profile and breadth), mATSI, PANAS, pre-activity, and SSQ. Each cell contains the mean and standard deviation.	76
Table 4	- Post-Activity Average Scores in order of participant completion: post-activity solar system questionnaire, PANAS, presence SUS, MSLQ, SAQ, BUZZ. Each score is listed with the mean and standard deviation.	77
Table 4 A	- Slater-Usoh-Steed Presence Questionnaire (Presence SUS) scores by group. Max value of 7; a higher score is better.	79
Table 5	- Average number of coded sections from each learner's simulation use by group for two code types: Movement Controls and Planet & Information Details. Note: these are individual code instances, not the total amount of session time described by that code.	80
Table 6	- Average number of coded sections related to rotating the view (PC) or turning physically (VR) per group.	80
Table 4 B	- PANAS post-use scores by group; higher scores represent high positive or negative affect.	82
Table 4 C	- SAQ subscales average scores by group; higher scores are better.	84
Table 7	- Average number of during-activity questions answered and extra facts discussed by learner.	85
Table 8	- Average number of pre- and post-activity solar system survey correct responses reported by group; higher scores are better.	86
Table 9	- Average number of pre- and post-activity solar system survey extra facts reported by group; higher scores are better.	86
Table 4 D	- MSLQ subscales average scores by group; higher scores are better, except for anxiety where a lower score is better.	88

Table 10	- Average number facts remembered and reported by learners during the follow-up interviews when asked to describe the experience or a detail or two about the planets.	91
Table 8 A	- Average number of post-activity survey correct responses reported by group; higher scores are better.	102
Table 11	- Number of facts remembered and shared during open-ended question in the follow-up interview.	103
Table 12	- Average UMUX score per group. Higher scores are better, with a max score of 24.	105
Table 13	- Average BUZZ scores per group. Higher scores are better, with a max overall score of 77. The Enjoyment Subscale is out of 35, and the Ease of Use Subscale is out of 42.	105
Table 14	- Forecasted and post-simulation use PANAS scores.	106
Table 15	- Average code counts for the four groups, representing the number of times that code was used to mark a section of the video. Note: these do not represent the total length of time the action/code marked quotations inside of each screen recordings.	108
Table 3 A	- Technology use Frequency Profile and Breadth scores by group; higher scores represent more use.	114
Table 3 B	- PANAS pre-simulation use scores by group; higher scores represent higher positive or negative affect.	116
Table 16	- Reported attendance of types of Informal Science Activities by group.	116
Table 17	- Reported Science/News Media (including TV shows like NASA, Discovery, or online news from Twitter, Facebook, etc.) following by group.	117
Table 3 C	- mATSI Subscale scores. Higher scores are better, except for anxiety subscale (lower scores are better). Total scores: Value of Science to Society = 25; Anxiety = {min = 5; max 25}; Self-Confidence in Science = 25; and Desire to Do Science = 35.	118

LIST OF FIGURES

Figure 1	Initial PC view of the solar system when started.	60
Figure 2	Small overview panel for Mercury, containing a summary of mass, diameter, density, temperature, and velocity.	61
Figure 3	Detailed overview panel for Venus shows four tabs: Overview, Motion, Composition, and Temperature, with numerical values for mass, density, and surface temperature showing.	61
Figure 4	Initial VR view of the solar system when started.	63
Figure 5	Vive Pro Controller.	63
Figure 6	Overview panel for Saturn shows on pointer hover.	64
Figure 7	Overview panel for Uranus shows on pointer hover, playing the density sound.	65
Figure 8	A204’s zoomed-in view of Jupiter showing the main moons, while answering other questions about the planet using the large information panel.	94
Figure 9	A204 pauses the simulation to view the two moons orbiting Mars.	94
Figure 10	C404 finding Uranus after teleporting to Oberon, searching to her left, and then below her.	96
Figure 11	D503 chasing down Uranus after teleporting to Oberon; she spent over 30 seconds following the orbits before pausing the simulation.	97
Figure 12	C401 looking up to search for a moon around Saturn.	98
Figure 13	C404 looking up to view the solar system from below.	99
Figure 14	D501 reflecting on the temperature differences between Mercury and Venus.	100
Figure 15	B302 exploring to find the coldest planets; the climate tab is open to play the temperature sonification.	111
Figure 16	D502 moving his hand up and down while listening to the mapping for length of day on Earth.	112

LIST OF SYMBOLS AND ABBREVIATIONS

BUZZ	Audio User Experience Scale
HCI	Human-Computer Interaction
mATSI	modified Attitudes Toward Science Inventory
MSLQ	Motivated Strategies for Learning Questionnaire
NASA TLX	NASA Task Load Index
NGSS	Next Generation Science Standards
PANAS	Positive and Negative Affect Schedule
PP	Planetary Perspective
presence SUS or SUS	Slater-Usoh Steed presence questionnaire
SEI	Student Engagement Instrument
SAQ	Science Activity Questionnaire
SSP	Solar System Perspective
SSQ	Sim Sickness Questionnaire
UMUX	Usability Metric for User Experience (Scale)
UX	User Experience
VR	Virtual Reality

SUMMARY

Schools use a variety of interactive software to support education, especially in STEM (science, technology, engineering, and mathematics) classes. Educational technologies provide a way to support 3D interaction and exploration for complex STEM topics, but many of them have not explored high-fidelity multimodal interactions. Previous studies have explored the best methods to measure emotional, cognitive, and physical engagement, but these methods have not been applied to fully understand the impact of multimodal interactive simulations on student learning. Technologies like Virtual Reality can provide a novel means for supporting interactive simulations for student learning. However, the full impact of these new systems and modalities on learning and engagement is unclear.

This study investigated different versions of interactive simulations for astronomy education. The dissertation included the design and evaluation of the sonification model for the solar system, which was then embedded within two different simulation versions. It evaluated a variety of tools for measuring and comparing user experience, engagement, affect, and learning, and compared qualitative differences between learner interaction in the four conditions. Other factors investigated included science anxiety, motivation, and technology experience, and their effect on a student’s ease of use and comfort in using newer technologies for education. The study found significant differences between the virtual reality (VR) and PC conditions and between the audio and no-audio conditions, with the VR and audio supporting better learning opportunities than the PC or no-audio conditions.

CHAPTER 1. INTRODUCTION AND MOTIVATION

1.1 Motivation

Lack of knowledge and interest in STEM has led the US to be ranked historically low compared to other countries: 38^{th} in math literacy and 24^{th} in science literacy (Desilver, 2017; Kuenzi, 2008). Even more recent measurements show a smaller disparity in science but still present a larger disparity in math (Andreas Schleicher, 2019; OECD, 2019). One attempt to mitigate this gap was the development of the Next Generation Science Standards (NGSS), a set of STEM standards that identified important concepts which are critical to any student's success in K-12 education (National Research Council, 2013). Included in these standards are core concepts within physical, life, and earth and space science, as well as application of these ideas in an engineering context.

One core idea at the middle school level is a broader understanding of space systems, including conceptual knowledge of our solar system, gravity, and interpreting size and physical properties of planets from data sources. Successful understanding of these topics stems from having an ability to interpret 3D models (Parker & Heywood, 1998). Through the Virtual Solar System Project, Barab et al. (2000) and Keating et al. (2002) have explored the usefulness of 3D computer modeling in a virtual solar system environment to support student understanding of moon phases and seasons. They found that the 3D models afforded visualization of abstract 3D concepts, but some of the students had an incomplete conceptual understanding of the phenomena.

This dissertation extended this initial work in 3D environments to study how a high-fidelity interactive, multimodal model of the solar system can impact learning and engagement outcomes for students at the middle school level. Virtual Reality provided a contextually situated environment where students can explore, control the scale, and build knowledge through their interaction. I expanded the feature set of an existing high-fidelity modeling system, Universe Sandbox (Giant Army, 2015), and investigated methods for effectively measuring the learning and engagement of students in this context.

1.2 Thesis Statement

As part of this work, my thesis contributed new knowledge on the best ways to measure multiple components of engagement for interactive educational technology and explored if they can be used as reliable measures compared to more typical measures of engagement. Through this work I identified which factors have a large impact on a student's ability to interact comfortably with different types of interactive simulations. My research will help the broader community better understand the impact of multimodal interactive systems, and whether or not VR environments have a greater impact (in the short or long term) on student learning and engagement.

1.3 Research Questions

The primary question addressed by my research is: How well can multimodal Virtual Reality systems support learning and engagement compared to typical interactive simulations for science education? To answer this research question, I studied the student experience with the multimodal tools:

RQ1. Does a VR simulation or PC simulation support higher levels of emotional, intellectual, and physical engagement?

RQ2. Does an audio-enhanced simulation or a visual-only simulation support higher levels of emotional, intellectual, and physical engagement?

RQ3. What factors, such as technology experience, math and science anxiety, self-efficacy, and affect, influence a student's ability to interact comfortably with multimodal science tools?

1.4 Summary of Studies

The first phase of research consisted of semi-structured interviews with five science teachers to identify a list of topics they typically teach, as well as the common misconceptions students across all levels struggle to understand. Some of these misconceptions included scale and size of the universe, seasons, and general knowledge about planetary characteristics.

Using information from the science teachers, guidelines from the NGSS, and a student misconception identifier study, I selected a set of information to convey through a sonification model of the solar system. Then, working together with sound designers, I developed a sonification model for conveying nine different data variables for planets in the solar system including length of year, length of day, mass, temperature range, gravitational strength, and type of planet (Tomlinson et al., 2017).

I conducted a preliminary user experience evaluation through a sonified planetarium show. This evaluation gathered feedback from the audience about the understandability and usefulness of the sonification model for interpreting the data about the planets (Tomlinson et al., 2017).

The next study evaluated the auditory display's ability to support learning. Here, I evaluated accuracy scores on a 10-item pre- versus post-test. Participants completed a listening activity in which they responded to different questions relating to the sonification model, including data interpretation questions requiring reflection on the model. A primary goal of this study was to create a reliable, valid user experience measure for auditory interfaces, correlating the overall outcome with a standard metric, UMUX (Finstad, 2010). I completed a factor and principal components analysis to design a set of standardized audio user experience questions (Tomlinson, Noah, et al., 2018). I evaluated the reliability of the scale as well, using Cronbach's alpha.

For the last study, I embedded the finalized sonification model inside of the Universe Sandbox (Giant Army, 2015) and completed a between-subjects lab study evaluating usability, learning, and user experience. During the study, screen and over-the-shoulder recordings were used to monitor students' interaction, and their differences were analyzed through qualitative coding. Other measurements of engagement (e.g., Science Activity Questionnaire (Meece et al., 1988) and the Slater-Usoh-Steed Presence scale (1994)) and affect (Watson et al., 1988) were collected to evaluate whether using the simulation changes learner perceptions about science and their overall knowledge about astronomy.

CHAPTER 2. BACKGROUND

2.1 Learning Theory

2.1.1 Learning

Learning has been defined as numerous phenomena, including the general process of expanding knowledge and understanding more details about a topic from a particular point of view (Greeno et al., 1996). Piaget's constructivist view of learning focuses on how children's knowledge grows over the course of their development; expanding on Piaget's work, Papert's theory of learning, constructionism, examines these knowledge structures through the types of activities a learner engages in, specifically viewing the learning through context (Ackermann, 2001). Constructionism is frequently simplified to a short definition of 'learning-by-making' (Papert & Harel, 1991), but Ackermann (2001) explains how constructionism helps us understand 'how ideas get formed and transformed when expressed through different media, when actualized in particular contexts' (p. 4).

Situated learning, or the grounding of knowledge in a set of socio-cultural experiences, provides one theory which expands Papert's view of constructionism, and allows for a more rich understanding of learning activities and the environments in which learning takes place (Collins et al., 1989; Dewey, 2007; Greeno et al., 1996; Lave & Wenger, 1991; Papert & Harel, 1991). Situated learning provides a way to help students build knowledge from meaningful activities and use authentic experiences to support better understanding (Brown, Collins, & Duguid, 2007). Brown, Collins, and Duguid (2007) describe how a student's perception of the learning activity is influenced by the tools and their use, and Kaptelinin and Nardi (2012) have discussed how Engeström's framework for Activity Theory can be used to appropriately analyze educational contexts due to its ability to deconstruct learning environments into their component parts, in order to better understand a student's interaction with those learning tools.

Problem-based learning is one approach which supports students engaging in authentic learning experiences (Blumenfeld et al., 1991) and can help students explore content in meaningful situations (Greeno et al., 1996). Similarly, inquiry-based learning provides another way to engage students in authentic and easily approachable methods, while paralleling learning outcomes similar to problem-based learning (Edelson et al., 1999). Both problem-based and inquiry-based learning leverage the unique characteristics of situated learning experiences to engage students. One difficulty with each of these is motivating students to continue engaging throughout the process (Blumenfeld et al., 1991), as the learning tasks can be more complex.

Instructional scaffolds provide three different strategies for engaging learners in these complex problems: providing visual representations to support initial understanding; allowing direct control and observation of the phenomena; and enabling learners to explore through multiple views (Quintana et al., 2004). Contrasting cases are another scaffold which can help learners analyze and interpret information; this technique has learners purposefully compare and contrast two (or more) examples and identify information which they may not otherwise notice and interpret (Barron et al., 1998). Both of these techniques, in addition to more traditional teacher-led scaffolding (Bransford et al., 2000; Vygotsky, 1978), provide means to support learner engagement in situated, user-controlled learning experiences.

2.1.2 Understanding and Learning in Context

Multiple factors affect a student's learning experience, ranging from their sociocultural background (Kaptelinin & Nardi, 2012) to their familiarity and experience with the educational technologies being used (Davies, 2011). Educational technologies can support a variety of situated learning experiences, but other characteristics may influence a student's comfort and ease of use when interacting with the technology, including how well it supports metacognition, their overall technology familiarity, their engagement with the materials, and their self-efficacy for learning.

2.1.2.1 Metacognition

One key goal of helping students learn while participating in these interactive, authentic activities is to support their explicit reflection on the learning. Metacognition refers to the process by which learners reflect on and try to modify and self-regulate their own learning experiences (Brown, Brewster, Ramloll, Yu, & Riedel, 2002; Flavell, 1979; Sperling, DuBois, Howard, & Staley, 2004). Contrasting cases have been found to support better metacognitive reflection compared to other methods when used in combination with educational technologies that scaffold the learning activities (Bransford & Schwartz, 1999; Vye et al., 1998).

2.1.2.2 Engagement

Fredricks, Blumenfeld, and Paris (2004) define engagement as being a complex construct which supports academic achievement and active participation in the learning experience. They identify the three components of engagement as behavioral, emotional, and cognitive engagement. Behavioral engagement is the student's willingness to be involved in academic activities (e.g., discussions, participation, etc.) in a socially responsible manner. Emotional engagement is defined by the learner's overall affective responses to the entire learning environment, including the materials, topics, and activities. Cognitive engagement encompasses the student's effort and willingness to participate in the learning experience.

Engagement is indeed a complex, inconsistently-defined measure, since it might include anything from different levels of subject-specific engagement to overall school engagement, and other behaviors such as effort, interest, motivation, or even time on task (Appleton et al., 2008; Jimerson et al., 2003). Appleton et al. (2008) explored the different conceptual definitions of engagement, and found that most generally, academic engagement is defined by the time on task and cognitive engagement includes the learner's perspective on autonomy in the learning environment and their personal value of learning. Shernoff and Csikszentmihalyi (2009) have defined engagement as high levels of enjoyment, interest, and concentration for learners, utilizing flow theory and its relationship to zones of proximal development (Shernoff et al., 2003; Vygotsky, 1978).

Some researchers have presented engagement as a measure expanding from the more typical definition of academic engagement to a more holistic curricular or institutional engagement, which instead measures the persistence in the interacting with the learning task through four factors: beliefs about knowing, cognitive tasks, self-efficacy, and stress (Bédard et al., 2012; Willis, 1992). In some cases, classroom participation was the biggest factor predicting achievement (and engagement), while others have found that perceived control over the learning activity might have a larger impact (Finn, 1993; Weiner, 1992).

Most work understanding the sub-component of behavioral engagement has focused on school-level studies (Alexander et al., 2016; Fredricks et al., 2004; Yazzie-Mintz, 2007) where they have found student autonomy has a positive effect on behavioral, cognitive, and emotional engagement. Skinner, Wellborn, and Connell (1990) found a correlation between perceived control and engagement, but argued that perceived autonomy (not being pressured to perform) and feeling socially connected to the teachers and students was more important.

2.1.2.3 Self-efficacy, Motivation, Anxiety, and Engagement

Pintrich and De Groot (1990) found links between a learner's value for achievement and their self-efficacy for the learning tasks had a positive effect on their cognitive engagement and their emotional engagement with the task outcome. Fortney (2016) found that student engagement increased as the learner had more control and increased self-efficacy, which promoted long-term learning persistence. Authentic tasks and activities have been found to support higher levels of cognitive engagement, and have led to better achievement outcomes for learners (Alexander et al., 2016; Nystrand & Gamoran, 1991). Motivation may be a big factor in continued persistent interaction with a learning activity, and can lead to higher levels of engagement (Eccles & Wigfield, 2002; Fortney, 2016; Ryan & Deci, 2000). Contextualization of learning materials have been found to support self-efficacy, to help maintain intrinsic motivation, and to provide extra determination for extrinsic motivations, which may lead to longer-term engagement with learning (Ryan & Deci, 2000).

Student motivation has been found to impact overall levels of cognitive engagement (Garcia & Pintrich, 1996; Pintrich & De Groot, 1990). To better understand the relationship between motivation and engagement, the Motivated Strategies for Learning Questionnaire (MSLQ), a set of self-reported Likert questions, has been developed (Garcia & Pintrich, 1991, 1996; Pintrich et al., 1991, 1993). The MSLQ is designed to be modular, with 50 questions probing learning strategies and 31 items measuring motivation, and 15 total modular sub-scales (Pintrich et al., 1993). While the initial validity and reliability of the scale were evaluated with college students, follow-up studies have found the subscales to be useful predictors of student motivation (self-efficacy, intrinsic values, and test anxiety) (Duncan & McKeachie, 2005; Pintrich et al., 1991, 1993; Pintrich & De Groot, 1990).

Understanding the overlap between academic anxiety and motivation, and their potential combined influence on engagement has been the focus of numerous scale validations (Gottfried, 1985; Harter, 1981). Gottfried (1982, 1985) found that academic intrinsic motivation can be positively related to competence on the materials, and that while some anxiety is useful as a motivator, too much can lead to students withdrawing from the learning process. More recently, Cahill, Gorski, and Le (2003) found that some amount of stress and urgency can help increase task focus for students. To better understand this relationship, Harter designed and validated a scale to measure a student's intrinsic and extrinsic orientation toward learning, and found that three of the items loaded onto an underlying motivational factor while the other two items fit as a cognitive-informational factor (Harter, 1981). Through correlation of the Children's Academic Anxiety Inventory

(CAAI) with the Children's Academic Intrinsic Motivation Inventory (CAIMI), Gottfried (1982, 1985) found that students who reported higher academic intrinsic motivation had higher school achievement, higher self-efficacy, lower anxiety, and lower extrinsic motivation for the learning activities.

2.1.2.4 Affect and Engagement

Appleton, Christenson, and Furlong (2008) and Jimerson, Campos, and Grief (2003) have both thoroughly explored the relationship between cognitive and affective engagement across two large reviews of engagement literature. In most instances, this measurement of affect related to the learners feelings toward their teachers, peers, and the school, but did not include their affect toward the learning materials or technologies (Appleton et al., 2008; Jimerson et al., 2003). Affect is a self-reported way to understand someone's mood and extract the impact of a set of materials or situation on it, for both positive and negative components (Watson et al., 1988). Watson, Clark, and Tellegen (1988) define positive affect as a subjective measure of alertness and enthusiasm, while negative affect is a subjective measure of distress and aversive moods such as fear and nervousness; they developed the Positive and Negative Affect Schedule (PANAS) to measure these two primary dimensions of mood. Forecasting affect may provide insight to someone's level of preconceived comfort before interacting with materials or technology, and could be compared to their actual experience, to understand the impact of those materials (Calderwood et al., 2016; Noah et al., 2016). Measurements of affect, and its impact on the learner's perception of the educational tools may be a predictor for their engagement.

More general measures of engagement have been explored (Appleton et al., 2006; Fortney, 2016; Fredricks & Mccolskey, 2012). Appleton and colleagues (2006), after completing one review of engagement literature, settled on a taxonomy for defining engagement which broke it into four total constructs: behavioral, academic, cognitive, and psychological. Finding a lack of work exploring cognitive and psychological engagement, they developed a scale to allows learners to self-report on those two engagement factors: the Student Engagement Instrument (SEI) (Appleton et al., 2006; Betts et al., 2010).

In addition to more general measures of engagement, motivation, and anxiety, specific measures of engagement have been developed to better understand student attitudes toward numerous subjects, including science and math (Gottfried, 1985). Multiple surveys exploring math and science interest, attitudes, and beliefs have been developed (Archer et al., 2016; Fortney, 2016; Meece et al., 1988; Miller, 1990; Post-Kammer & Smith, 1985; Weinburgh & Steele, 2000). Engagement measures can provide another lens to examine a student's potential learning outcomes (short and long-term), but Forney (2016) found that one of the best predictors for a student's success in science is based off of their previous experiences in science classes. The Colorado Learning Attitudes about Science Survey (CLASS) was developed to understand student perceptions and beliefs about physics, and has been adapted to fit other disciplines as well (Adams et al., 2005, 2006).

Understanding potential interactions between socio-cultural variables, familiarity with technologies used in the classroom (especially those used in science education), and engagement is a broader goal which has not been well-explored (Jimerson et al., 2003). Some have begun addressing these other factors by studying how socio-economic factors, masculinity and femininity, and ethnicity impact a student's engagement with science in formal and informal contexts (Archer et al., 2016; Post-Kammer & Smith, 1985; Weinburgh & Steele, 2000).

Weinburgh and Steele (2000) designed the modified Attitudes Toward Science Inventory (mATSI) as a measurement of anxiety, self-efficacy, and personal beliefs about science, and evaluated its reliability within different sub-groups of gender, grade level, and ethnicity. Meece, Blumenfeld, and Hoyle's (1988) Science Activity Questionnaire (SAQ) was developed to measure cognitive engagement and goal orientation for learning activities. They found that students who had a goal orientation of task mastery (instead of social recognition or work-avoidance) reported higher levels of cognitive engagement (Meece et al., 1988), and Miller (1990) found that students participating in hands-on learning activities have higher SAQ cognitive engagement scores.

Engagement is a complex construct which includes cognitive, behavioral, and emotional components, each of which can be influenced by affect, motivation, anxiety, self-efficacy, previous experiences in specific domains, and the situated learning experiences. More comprehensive understanding about the socio-cultural influences, the student's metacognitive skills, and their familiarity and comfort using the educational materials could help provide a more complete view of a student's engagement with learning.

2.1.3 Multimedia Learning

Other factors which may affect learning include the presentation medium of the learning materials. Baddeley's model of working memory proposes sensory modality differences, or changes in comprehension and reasoning based on the channel processing the content (meaning visual or audio channels are processed separately) (1992). Another commonly accepted model, Paivio's Dual Coding Theory, instead focuses on the differences between verbal and non-verbal processing (Clark & Paivio, 1991). Mayer, more similar to the complex system presented by Wickens in Multiple Resource Theory (2008), instead proposed the principle of multimedia learning: presenting material using different modalities (e.g., words and pictures) in order to take advantage of differences in cognitive processing for sensory modalities *and* presentation modes (2001a). This theory about the cognitive processing of multimedia learning proposes different configurations of resources are used to process images and speech, while representations like text use a combination of visual and speech resources.

The multimedia principle argues that students learn better from a combination of images and text rather than text-only, and includes nine different principles for successful multimedia design (Mayer, 2002a). When utilized properly, previous research has shown the multimedia principle can successfully reduce cognitive load for learners (Mayer & Moreno, 2003). Schweppe and Rummer have found that written text and animations led to better transfer performance when tested after a delay, compared to narration and animation (2016). Multimedia tools have persistent, longer-term learning outcomes (Schweppe et al., 2015), and have been found to be effective when employed in classroom contexts, particularly for shorter intervention sessions (Harskamp et al., 2007). There is still discussion about the overall effectiveness of multimedia at supporting better learning outcomes than typical interventions. For example, Austin reported how text position and unnecessary animation can negatively impact learning, and emphasized the need for future research to study optimal display design characteristics for educational technology resources (2009).

The multimedia effect has been studied across multiple educational context, and typically relies on an intervention which compares text, pictures, narration, and other types of visual or speech representations with varying levels of detail (Dubois & Vial, 2001; Harskamp et al., 2007; S. H. Liu et al., 2009; Witteman & Segers, 2010). Leutner highlights three areas where recent work had additionally studied the impact of multimedia learning design decisions on interest, motivation, and emotional engagement, instead of focusing on the purely cognitive impacts (2014). However, even with its successes in supporting learning of difficult topics (Mayer, 2001b), there is a significant dearth of studies evaluating additional modalities, including tangible, tactile, and particularly relevant to this dissertation, non-speech audio.

2.2 Educational Technology

Educational technology has been explored as a way to support situated learning experiences and promote scaffolding (Blumenfeld et al., 1991; Edelson et al., 1999; Guzdial, 1993). In addition to supporting topic-specific learning, scaffolding could more broadly help learners strengthen their metacognitive practices, which can help them participate in inquiry-based learning activities which may lead to deeper cognitive engagement (Greeno et al., 1996; Quintana et al., 2004). These technologies can also support modeling through interactive simulations, which can help learners understand and engage in conceptual change, especially in the sciences (Nersessian, 1992, 1999).

2.2.1 Interactive Simulations

Interactive simulations are typically defined as technology-based environments which have underlying models and logic, oftentimes built to represent a natural real-world phenomena (D'Angelo et al., 2014). Simulations can also be categorized as virtual laboratories (mimicking typical lab experiments) or simulations of scientific phenomena (models which support observation of otherwise-difficult phenomena) (Scalise et al., 2011). Through dynamic modelling, these tools can help students recognize patterns and interact with complex scientific phenomena (Dede, 2000)

A large amount of work has explored how well simulations can support scientific inquiry and their overall impact on learning experiences for students. D'Angelo and colleagues' recent meta-analysis included 59 studies where interactive simulations were compared against similar instructional content; they found a strong effect size between the simulation and control group, with the simulation group having a larger percentage increase compared to the control (2014). Additionally, they found that a smaller subset of research has explored the impact of simulations on scientific reasoning, inquiry, and non-cognitive outcomes (though fewer studies explored these). Others have specifically reviewed simulation reach at the middle and high school levels (Scalise et al., 2011) and for college (Ma & Nickerson, 2006a). Across these reviews, it has been found that assessments of these interactive simulations do not regularly complete evaluation in the context of the virtual environments, but instead rely on surveys and open-ended questionnaires for assessment (D'Angelo et al., 2014; Ma & Nickerson, 2006b; Scalise et al., 2011).

Games and immersive simulations can help situate knowledge and support inquiry-based learning for students (Barab & Dede, 2007). Using authentic materials, realistic situations, and models of those events (simulations) can provide enhanced transfer of information and skills from one task to another (Halpern, 1998). Having students reflect on these learning opportunities can also support better metacognitive reflection and attempt to correct faulty logic (Halpern, 1998). Numerous projects have explored the ability for simulations to support these authentic learning experiences. The Design Principles Framework has noted basic, intermediate, and advanced themes across 79 different studies examining patterns for simulations across interface design (including scaffolding), visualization (e.g., perspective shifts, control of speed, and zooming), and scientific inquiry (i.e., how they support data gathering) (Scalise et al., 2011).

2.2.2 Virtual Reality (VR)

Virtual reality is a virtual environment which, instead of overlaying and enhancing the normal world, instead replaces it. Steuer defines VR as an experience, mediated by technology, where one feels surrounded by an environment which is not the immediate physical world (1992). Heeter identified three dimensions which affect presence, including subjective personal presence (feeling embedded in the virtual world), social presence (extent to which you feel connected to others), and environmental presence (how much the environment reacts to your actions) (1992). While the fidelity of these virtual environments has changed over time, presence still significantly effect experiences with VR (McGlynn & Rogers, 2017). These experiences are typically supported through HMDs like the Oculus Rift or the HTC Vive, in an attempt to completely embed the user into a virtual world. VR systems are usually defined by the technological capabilities and not the overall experience from the user's perspective (Slater & Sanchez-Vives, 2016; Steuer, 1992).

2.2.2.1 Human-Computer Interaction of Virtual Reality

VR allows for an immersive visualization which can be highly contextualized, and they typically support control and manipulation of the virtual environment. One important characteristic of measuring the level of immersion in VR is presence. Spatial presence has been broadly explored by researchers in the field of VR, but Lessiter, Freeman, Keogh, and Davidoff argue that though spatial presence applies for both real and virtual experiences, many people do not consider how 'spatially present' they are in a real-world environment (Lessiter et al., 1998). This concept provides the possibility of people comparing their perceived presence in virtual experiences to their real-world experiences (Lessiter et al., 1998).

Immersion is the technical ability of a system to allow a user to perceive within that virtual environment through their own natural sensorimotor abilities, while the subjective measure which correlates to that amount of immersion is defined by presence (Slater & Sanchez-Vives, 2016). Immersion has also been identified as having four properties, being inclusive (blocking reality), extensive (supporting a range of sensorimotor activities), surrounding (panoramic vs. narrow field of view), and vivid (fidelity and richness of the environment) (Slater & Wilbur, 1997). The combination of immersion and presence may be driven specifically by the place illusion (feeling that you are in a specific location while you know that you are not) and the plausibility illusion (feeling that events can and are really happening) (Slater, 2009).

Understanding the relationship between presence and physical engagement for VR systems plays a significant role in evaluating them. Users of the VR system who have a higher presence rating than others may perceive the environment as being more engaging and may perform better on a task within that environment (Slater & Wilbur, 1997).

Discussion about the standard ways to measure presence in virtual environments is ongoing (Hofer et al., 2020; Lessiter et al., 1998; Slater, 1999; Usoh et al., 2000; Witmer & Singer, 1998). Lessiter et al. (1998) designed and evaluated the ITC – Sense of Presence Inventory (ITC – SOPI), where 44 items grouped to measuring four factors: sense of physical space (how 'located' someone feels within that environment), engagement (how involved and interested they are), ecological validity (how believable is the VR environment), and negative effects (how much simulation sickness, headaches, or other effects may be reducing someone's experience. Witmer and Singer developed two separate questionnaires, the 32-item presence questionnaire (PQ) and the 29-item immersive tendencies questionnaire (ITQ) (Witmer & Singer, 1998). The PQ measures someone's level of presence for a particular VR system, while the ITQ provides a more general measurement of how likely someone is to become immersed within a system (Witmer & Singer, 1998).

In response to those longer surveys, Usoh, Catena, Arman, and Slater designed a set of six Likert questions (and one free response), called the SUS (Slater, Usoh, and Steed; or, here the 'presence SUS'), meant to evaluate whether or not someone feels present in a real or virtual environment (Slater, 1999; Usoh et al., 2000). Instead, the presence SUS may be more helpful for measuring a sense of presence across similar media (both virtual environments) and not between virtual and reality (Slater et al., 1994; Usoh et al., 2000).

More recently, Hofer et al. (2020) also used a shorter eight-item Spatial Presence Experience Scale (Hartmann et al., 2016) and a three item questionnaire to evaluate participants' perceived spatial presence and plausibility while in a virtual environment. Measuring the immersion and presence of a user in a VR environment is important to understanding possible differences in engagement and learning; one possibility for exploring this at a higher level is McGlynn's Magnet Model of Spatial Presence (MMSP), which explores the differences between physical and virtual stimuli and their effects on presence (McGlynn & Rogers, 2017).

2.2.2.2 Audio in Virtual Reality

Audio embedded within a VR environment has been explored a bit, but mostly as a way to increase someone's presence within that system (Bormann, 2005). Bormann found that spatialized audio in a virtual environment can affect both task performance and can lead to higher levels of presence for a system (Bormann, 2005). Increased fidelity of audio in VR can be used for educational applications (Dalgarno & Lee, 2010) and has been used to support better presence in video games (Lachlan & Krcmar, 2011).

Even in more popular mixed reality systems, the use of auditory interfaces remain largely unexplored (Billinghurst & Kato, 2002). One notable exception is Chatzidimitris, Gavalas, and Michael's SoundPacman, where they studied the ability for 3D audio in mixed reality environments to support game immersiveness (Chatzidimitris et al., 2016). Others have integrated audio into mobile games, usually as a way to support immersion and emotional engagement (Paterson et al., 2010). Embedded, spatialized audio to promote location awareness (typically through the use of Head Related Transfer Functions, or HRTFs) has been used more across mobile AR platforms (Ekman et al., 2005; Pellerin & Bouillot, 2009).

Understanding how people complete localization for 3D sources within the virtual environment helps to drive design guidelines for using spatial audio into future applications. There are a few examples of audio-enhanced musical experiences where a user can directly place instruments or sound sources are placed in the 3D space around themselves (Haller et al., 2002). Sodnik and colleagues found that HRTFs can provide satisfactory localization cues in AR when visual cues are present for the virtual objects (2004). Further work by Vazquez-Alvarez et al. (2015) explored the differences between spatialized audio in exocentric (a fixed view) and egocentric (from the user's perspective) displays for a multi-level auditory display supporting an interactive art installation. They found that both types of displays using spatialized audio cues supported longer exploration, but using the same type of display for the primary and secondary spatialized audio led to higher cognitive load for interpretation (Vazquez-Alvarez et al., 2015). In general, people can localize spatialized audio in virtual environments, with relatively low cognitive load (particularly when representing different content), and it can provide meaningful exploration in an interactive system.

2.2.3 Auditory Displays

Auditory displays are displays which use purposefully designed sound to convey information. Sonifications are non-speech auditory displays which use data-driven mappings to support interpretation or comprehension of the meaning (Kramer, 1994; Walker & Nees, 2011). A variety of auditory display types exist, including auditory icons, earcons, audification (more direct mappings between the data and representation, e.g., speeding up playback rate of seismic waves to make them audible), and sonification (Brewster et al., 1993; Gaver, 1986; Walker & Nees, 2011). Auditory icons are realistic sounds, such as a door opening or shutting, which may have a direct or associated meaning (Gaver, 1986). Earcons are usually musical or synthetic sounds with learned representations and associations (Brewster et al., 1993). Auditory icons are usually faster for someone to learn, but are restricted by the number of real-world associations which could be integrated into a display; earcons may take more practice for a user to gain familiarity (Dingler et al., 2008; Kramer et al., 1999). For example, an auditory icon for rain would have a faster recognition and interpretation because of its real-world association, but an earcon representing cloudy may not have the same initial association.

While some auditory displays focus on one variable, or stream of information, Bregman's comprehensive work in auditory stream analysis examines the process by which people perceive multiple streams of auditory information and interpret them individually or holistically (1990). More complex sonification work has explored the design and evaluation of multiple streams, especially in applied data interpretation contexts (Brown et al., 2002; Schuett, 2015; Schuett & Walker, 2013).

2.2.3.1 Auditory Displays in Education

Sonifications can leverage metaphors, previously associated meanings (especially from auditory icons), and easily interpreted mappings to support initial understanding. Some common metaphors include the temperature to pitch (perceptual interpretation of the sound frequency) or size to tempo mappings (where smaller objects are 'faster' than larger objects) (Flowers, 2005; Walker & Kramer, 2005).

In addition to completing analysis or exploration of a dataset using sonification, other research has found its usefulness in supporting pattern-finding, exploration, point estimation, and in interpreting uncertainty and error (Batterman & Walker, 2012; D. R. Smith & Walker, 2005, 2002). Flowers (2005) found that sonifications can improve recognition and recall for students, especially when using careful sonic information design. Upson (2001, 2002) found that sonifications can lead to greater engagement for students when learning about Cartesian graphing concepts.

Auditory graphs and sonifications can also provide means for students with vision impairment to explore graphs through audio (Brown et al., 2002; Mansur, Blattner, & Joy, 1985; Stevens, Brewster, Wright, & Edwards, 1994; Tomlinson, Batterman, Chew, Henry, & Walker, 2016; Upson, 2001). Auditory graphs can be more flexible than many of the tactile education tools, as they are easier to adapt and require less physical equipment to build compared to many of the physical graphing tools.

2.3 Educational Technology Evaluations

2.3.1 Learning Studies in Interactive Simulations

Evaluations of interactive simulations have ranged from small-scale surveys and interviews after embedding those materials into coursework (Perkins et al., 2006) to longitudinal assessments over the course of a semester or year (Barab et al., 2000; Keller et al., 2006; Rehn et al., 2013). One meta-analysis of simulation evaluation found there are three general types of evaluations for interactive simulations: impact on achievement outcomes, ability to support scientific inquiry and reasoning, and non-cognitive outcomes (D’Angelo et al., 2014). For example, for the PhET simulations (PhET sims, or simply sims), evaluations have spanned screen and audio recordings of students using the simulations, as well as field notes, and individual interviews with students (Finkelstein et al., 2005; Rehn et al., 2013; Tomlinson, Batterman, et al., 2018; Tomlinson, Kaini, et al., 2019) and other learners (Tomlinson, Walker, et al., 2020).

In the two case studies outlined by Rehn et al. (2013), recordings were analyzed to better understand how well the simulations met the heuristics for simulation use (i.e., how well the simulations and their instructional materials supported desired behaviors and learning outcomes). One evaluation of PhET sims included hundreds of individual interviews with students to observe how they encourage students to complete ‘engaged exploration’ while using them (Adams et al., 2008). Other evaluations have used guided activity observations and post-activity performance measures to evaluate how well the simulations support student exploration and learning for activities typically restricted to physical interactives (Finkelstein et al., 2005). In this case, it was found that students who used the interactive simulations had better performance on conceptual questions and better skills in building circuits compared to learners in the other conditions.

Other evaluations of interactive simulations have begun to explore additional measures (e.g., eye tracking) for tracking which pieces of the simulation students are using. She and Chen (2009) studied how multimedia tools can help middle schoolers learn about mitosis and meiosis in cells by evaluating post-test and retention-test scores, as well as comparing mean fixation time during the learning activity. They found that students in the simulation group with additional on-screen text information performed better on post and retention tests than the simulation group with narration and the animation group with on-screen text.

Barab, Hay, Barnett, and Keating (2000) evaluated the impact of interactive simulations in a slightly different context; instead of having students only explore a model, they completed a two-year evaluation where the tool was used by students in a project-based context to build their own models of the Earth/Moon/Sun system to learn about physics and astronomy concepts in a contextualized way. During this classroom deployment, they taped student groups working on the project, conducted interviews with students and teachers, and completed retrospective analysis of the videos and artifacts (Barab et al., 2000).

In this work, they found that the 3D simulations helped students learn in three ways: 1) they provided an authentic context to explore the information (e.g., situating the problem into a realistic problem space); 2) they supported metacognitive practices (e.g., being able to use the model to reflect on what they do and do not know when thinking through the problem set); and, 3) they supported discourse between and within groups of students (e.g., the students began to adopt astronomy terminology as they progressed over the course of the project) (Barab et al., 2000). The Virtual Solar System project explored the potential of interactive simulations being used in a situated, project-based context, and found that students successfully learned astronomy concepts while engaging in this learning activity.

2.3.2 Learning Studies in Virtual Reality

VR has been repeatedly used as a way to support subject-specific applications for learning (Keating et al., 2002; Winn et al., 2002). VR allows learners to interact with different environments and situations, no matter the size, scale, or level of complexity represented by that model, supporting a large variety of contexts and applications

(Christou, 2010). VR can support multimodal experiences within a comprehensive environment, leveraging the ability for systems to provide visual and audio representations, while using embodied interaction and movement to situate the learner (Christou, 2010).

Slater and Sanchez-Vives studied how VR environments can support tangible exploration of abstract concepts (through control, rather than static observation) and found it to be helpful for supporting scientific visualization and comparison of macro and micro scale scientific phenomena, especially in cases where students can break outside of the realistic constraints for that problem (i.e., allowing them to visualize things which cannot be done in real life) (Slater & Sanchez-Vives, 2016).

The ability for virtual environments to help students investigate spatial relationships has been extensively studied (Keating et al., 2002; Shelton & Hedley, 2002, 2004). Shelton and Hedley completed a deployment of an augmented reality system for helping students learn about astronomy relationships between the Earth and the Sun (e.g., seasons, revolution, and equinox/solstice) (Shelton & Hedley, 2002). During this deployment, 34 students used different experiences meant to highlight these concepts. They completed a pre-post evaluation of the students' knowledge and transcribed the interviews from each session. Through this work they found that augmented reality provided a flexible way for students to explore a scaled representation, provided the ability to explore time changes, positions, and angles, allowing them to build procedural, declarative, and configurational knowledge in the spatial visualization. Students in the augmented conditions had better performance in interpreting 3D geographic visualizations and had more accurate understanding of spatial tasks than students in a desktop condition (Shelton & Hedley, 2002, 2004). More recently, Edwards et al. evaluated a VR system to support learners in understanding visual spatial content related to organic chemistry, and reported high participant perception of usefulness for instruction and motivation (2019).

Overall, VR has been found to provide an interactive and meaningful environment to visualize complex and abstract concepts (Saidin et al., 2015; Wang et al., 2017). Wang and colleagues (2017) provide an overview of their recent work on evaluating AR in education contexts, and have generally found that these systems increase enjoyment, promote self-driven learning, and facilitate development of communication skills (when used collaboratively). In a metareview of 167 studied Reisoğlu and colleagues (2017) found that many evaluations of 3D educational environments study presence, self-efficacy (emotional and behavioral factors), and some study cognitive achievement, though there were not many studies which explored a combination of these factors in one evaluation, and many studies took place in the platforms of Second Life or Active Worlds. Examining the impact of VR environments across multiple factors, including cognitive, behavioral, and emotional ones, is important to understand whether or not they are effective learning tools schools should use.

2.3.2.1 Embodiment

Embodied interactions through touch screens and other educational technologies lets the learners explore, interact, and use other sensiomotor and attentional resources which they might not use during a passive learning experiences, leading to more engagement (Abrahamson & Sánchez-García, 2016). Physical interaction has also been found to increase on-task performance for students (Mahar et al., 2006) and younger students who have some form of physical activity throughout the day have higher reported classroom behavioral scores (Barros et al., 2009).

Embodiment can help students move from using technology to support actions to using technology to support thinking, where physical interaction with tools can support reflection and metacognition (Abrahamson & Sánchez-García, 2016). Goldin-Meadow and Beilock (2010) suggest that gestures can affect thinking through the internalization of an action, and that thinking through the gesture can help someone solidify their mental representation of a concept. Hall and Nemirovsky (2012) agree with this sentiment, and argue that embodied cognition supports problems solving but discuss how gestures help represent this embodiment through sensorimotor interaction.

Lindgren and Johnson-Glenberg (2013) propose six guidelines for using VR for education; one guideline refers to building action-concept relationships, and outlines how these elicit and affect physical activity in that environment, especially through immersion and congruency between the gesture (action) and the content area. Dawley and Dede (2014) found that because of embodiment, VR and other mixed reality learning environments can exploration and experiential learning (instead of a recall-only experience). One example of this found that students who were more physically active in the virtual learning environment had a greater change in understanding for spatial relationships (Shelton & Hedley, 2004). VR provides a good context for helping students take advantage of embodied learning, which might lead to better learning outcomes and more motivation while engaging in the educational materials.

2.3.2.2 Quantitative Studies

Generally, many studies of VR educational technologies have not used quantitative evaluation methods. Wu, Lee, Chang, and Liang (2013) discuss the potential for virtual environments to support more authentic, task-based exploration approaches which could lead to new ways to measure engagement and learning for students, though it has not been studied in a rigorous and generalizable manner. Billinghurst and Dünser (2012) provide an overview of some of their studies which have been done comparing an virtual environment to more traditional classroom materials, and only one used a quantitative measure (a post test and a follow-up retention test), while the others relied mostly on qualitative information. Even the quantitative evaluation from Edwards et al. included only self-reported, individual questions about motivation, ability to support multisensory learning, haptics, usefulness as an instructional tool, and overall experience (2019). While qualitative feedback can help situate the information gathered about experience differences between the two systems, it does not provide the whole picture and understanding which quantitative methods could be used to compare virtually immersive learning experience with a more traditional one is important for completing comprehensive evaluations.

2.3.2.3 Qualitative Studies

The majority of VR studies rely on qualitative coding from videos to understand the effect of the technology of student levels of learning and engagement. In one classroom deployment, Kerawalla, Luckin, Seljeflot, and Woolard (2006) completed qualitative coding for three types of data: classroom recordings of the teacher and students during the virtual environment-supported lesson, the recordings from the traditional lessons, and transcripts from follow-up interviews. The follow-up interviews provided some reflection about successes and failures about integrating the technology into a lesson, and student engagement during the activities, but did not complete this in any structured manner. Qualitative analysis is typical for interviews and classroom deployments; however, it is possible that interviews with teachers (regarding a student's engagement with the material) may provide a biased view based on that teacher's previous ideas about how a student participates in classroom activities. Combining both qualitative and quantitative methods can help provide a clearer understanding about a student's overall experience using an educational technology.

2.3.3 Human-Computer Interaction (HCI) Evaluations

Evaluation methods from HCI may provide another lens to analyze a student's experience interacting with educational technology, especially when the comparison may be taking place between a typical technology (a computer simulation) and a newer modality, like VR.

2.3.3.1 Usability and User Experience

Jakob Nielsen defines usability as an 'quality attribute that assesses how easy user interfaces are to use, including characteristics such as learnability, efficiency, and satisfaction' (Nielsen, 2012). Best practices for measuring the usability of a system are an ongoing discussion in the Human-Computer Interaction and User Experience communities. Many scales have been developed and validated to measure the usability, including the Usability Metric for User Experience (UMUX), the System Usability Scale (SUS), and the UMUX-Lite (Brooke, 1996; Finstad, 2010; Lewis et al., 2013).

The SUS is a 10-question scale designed to measure the efficiency, effectiveness, and satisfaction of a system, paralleling the ISO standard 9241-11 (Brooke, 1996, 2011, 2013). The scale is made up of 5-point Likert questions, where respondents state how much they agree or disagree with the statements. The UMUX is a 4-question scale which has been presented as a shorter alternative to the SUS; it has been found to highly correlate with SUS scores (r = 0.96) and to be highly reliable, with a Crombach's alpha of 0.94 (Finstad, 2010, 2013). While these scales provide quantitative measurements for usability, it can be difficult to have a clear interpretation of how well that score represents the overall User Experience. Bangor, Kortum, and Miller (2009) found that adding an 11th question to the end of the SUS asking for a verbal description about the user-friendliness of the system can help interpret the quantitative scale information in a meaningful way.

While these scales can provide a baseline evaluation for UX, it is difficult to use them for evaluating auditory displays. Most people have low experience using auditory displays, except for navigation-based text-to-speech ones (e.g., Google Maps), and so interpreting usability scales for evaluating these displays can be confusing. There is a need for other measurements which could help evaluate the ease of use and aesthetic appeal for these auditory displays.

2.3.3.2 Technology Familiarity

A student's overall comfort level and familiarity with technology could also have an impact on how well a student interacts with a new AR or VR system. If they do not have previous experience with that technology, or something like it, they may feel more uncomfortable and less likely to explore freely in that system. Exposure to different types of technology at school, at home, and in other locations (museums, friend's houses, etc.) may have an effect on a student's emotional and cognitive engagement with the technology or educational topics.

Beer created a Technology Experience Profile to understand the frequency with which someone (older adults in this case) may use a set of technologies, including ones for communication, recreation, computer and mobile, and transportation (2013). This Technology Experience Profile has been adapted by Karina Liles to help understand with which newer technologies students are familiar (2017). While student familiarity with technology may impact their overall comfort and affect exploring a new environment, an in-depth understanding of its impact on students has not been completed at any large level.

2.4 Educational Topic Focus

2.4.1 Next Generation Science Standards (NGSS)

The Next Generation Science Standards are a set of science, technology, engineering, and mathematics (STEM) standards designed to specify important concepts which are critical to student success at the K-12 level (Achieve, 2013; National Research Council, 2013). These standards include concepts such as physical, life, earth, and space sciences, and outline the application of these ideas to engineering context.

One area outlined in the NGSS for the middle school level in the Earth and Space Sciences category is ESS1: Earth's Place in the Universe (Achieve, 2013; National Research Council, 2013). There are three main goals for ESS1, to have students be able to show understanding by being able to:

> ESS1-1. Develop and use a model of the Earth-sun-moon system to describe the cyclic patterns of lunar phases, eclipses of the sun and moon, and seasons.

ESS1-2. Develop and use a model to describe the role of gravity in the motions within galaxies and the solar system.

MS-ESS1-3. Analyze and interpret data to determine scale properties of objects in the solar system.

Portions of these three learning outcomes for MS-ESS1 will be the focus of the interactive simulation deployment in this work.

2.4.2 Common Core

For an additional comparison, many of the NGSS have connections to Common Core standards in literacy and mathematics. The middle school literacy standard (RST.6-8.1) looks for evidence that a student can explore science and technical information to support analysis and argumentation (National Governors Association, 2010). A second literacy standard (RST.6-8.7) wants to see that a student is capable of interpreting quantitative or subject-specific information from diagrams, models, or graphs (National Governors Association, 2010).

The Common Core mathematics standards 6.RP.A.1, 6.RP.A.2, 6.EE.B.6, 7.EE.B.6, and MP.4 all outline how students should be able to understand models that use math, to interpret ratios, and to understand how to interpret variables which represent different quantities in an inequality or system of equations (National Governors Association Center for Best Practices, 2010). The overlap of the NGSS with the Common Core standards provides an interesting, important, and unexplored area to situate this work.

2.4.3 Space Simulations

One freely available set of simulations built to support learning about these concepts is from PhET Interactive Simulations (PhET Interactive Simulations, n.d.). PhET sims use the principle of implicit scaffolding to support students freely exploring the sims and trying different scenarios to learn about the underlying goals of the simulation (Moore et al., 2014; Perkins et al., 2006). Gravity and Orbits is one PhET sim which focuses on astronomy concepts mentioned here (PhET Interactive Simulations, 2018). Other free, JavaScript-based space simulations exist, including Vezenia's WebGL Solar System & orbital mechanics simulator (Vezina, 2017) and the Sky Live's 3D Solar System Simulator (The Sky Live, 2018). A few desktop and mobile-based simulators are also available from NASA and Inove (Inove, 2017; NASA, 2015).

The Universe Sandbox is a more complex space simulator, more similar to the NASA and Inove developed tools, which supports open-ended scenarios such as solar system modeling, introduction of additional bodies to a system, realistic gravity for large-scale space collisions, and complex star systems (Giant Army, 2015). Giant Army provided access to their desktop and virtual reality code in order to support this work, and the designed audio model for the solar system was integrated into their program for exploration and testing.

CHAPTER 3. PRELIMINARY WORK

3.1 Study 1: Planetarium Study

Museums and other informal learning environments (ILEs) have been exploring the use of multimodal (e.g., visual, audio, tactile) exhibits to increase engagement, prolong interaction with the educational materials, and support shared experiences for those with vision impairment (Allen, 2004; Horn et al., 2009; Walker et al., 2006). In one example, the Aquarium Sonification project mapped tank events and presented individual fish characteristics through dynamic soundscapes as a way to build a unique experience leveraging interaction with auditory displays (Jeon et al., 2012).

Similar to other ILEs, planetariums use visual resources like diagrams, pictures, and videos or animations, paired with detailed verbal description to convey information about space. The same types of materials are used by instructors when covering astronomy concepts in a formal learning environment, too. Through the development of a planetarium sonification, I was interested in exploring how we could leverage previous work in ILEs and the spatial audio afforded by the planetarium's physical set-up to convey a variety of quantitative information about each planet. Designing evaluation materials which could help understand an audience member's listening experience (i.e., their user experience of interpreting the auditory display), their overall engagement (i.e., their affect and the influences of the aesthetic design), and the amount of information they learned or understood in a different way (i.e., learning which occurred due to their continued engagement and interpretation of the display) was another goal of this work.

3.1.1 Teacher Interviews

Initial work on the development of the planetarium study meant deciding which astronomy concepts should be highlighted and what details should be incorporated into the show. To generate a better understanding of what topics to include, we conducted semi-structured interviews with five different astronomy teachers (three women): one elementary science teacher; two instructors, from the Center for Education Integrating Science, Mathematics, and Computing (CEISMC), who taught at the middle and high school levels; one college astronomy professor; and a planetarium instructor for a local science center (Fernbank). We recruited three teachers by reaching out through CEISMC, and two though direct email. Each interview was about one hour long and took place at their private office or classroom, outside of their teaching hours.

Before the interviews, each teacher completed a demographics survey where they self-reported information about their total number of years teaching, their educational background (on science and math topics), general teaching experience, and types of activities they use to engage students in the material during class (see Appendix A for these questions). Each instructor had at least nine years teaching, and many of them had over 20 years of experience teaching or working on curriculum development for science and math.

Each interview was audio recorded and additional notes were recorded in the interview guides by both the interviewer and note taker (see Appendix A for these questions). As part of the interview, we asked the instructors to report the types of material they introduce in a typical lesson, what activities they use to scaffold introduction of these concepts, and if there are any common misconceptions for their students (or audience).

These interviews provided a means to explore how the topics in the planetarium sonification should be introduced and to understand what an appropriate level of detail was.

3.1.1.1 Interview Results

One common theme which arose was the general lack of knowledge that the public (students and adults alike) have about astronomy: teachers typically need to start a lesson (or even a planetarium show) by introducing the terminology which will be used throughout that activity. When introducing the topics, some teachers probed students more deeply to find out what type of baseline knowledge students might have (e.g., about constellations, planets, the moon, and orbits). Much of astronomy knowledge is detail-specific; to make it easier for students, many teachers described comparing and contrasting features of the planets such as surface composition (planet type), rings, moons, and other planetary details. Since the planetarium instructor may have an audience with more diverse knowledge, she reported trying to mix different levels of details into the show (to supplement the basic details the audience might already know). The other instructors typically had more homogeneous groups of learners, so it was not such a major concern to them.

When asked to report common misconceptions, many of the teachers identified comprehension of the size and scale of the solar system, the cause for changing seasons and phases of the moon, and the understanding orbits to be major topics of misunderstanding for students. All of the teachers discussed different activities they used to address these misconceptions; these activities ranged from using simulations or performing a physical activity to explain the underlying concept, such as modeling the phases of the moon and having students reflect on what they thought might happen compared to what actually happens. Labs, demonstrations, and physical activities provide interactive ways for students to critically reflect on their current knowledge and provide a more meaningful way to modify their misconceptions.

3.1.1.2 Reflection on the Interviews

One goal of these interviews was to identify the topics which teachers cover and the level of detail which is appropriate for a general audience. Though each instructor reported following different standards (including the Regents questions, the Next Generation Science Standards, and the Space Exploration AP Exam) for which details they covered, there was a large amount of consistency in the types of information each of the teachers described. From these interviews, it was clear that the details which should be included in the planetarium study should relate to the size and scale of the solar system, and other comparable details between planets (e.g., mass, temperature, distance from the sun) (Tomlinson et al., 2017).

3.1.2 Misconception Identifier Study

After completing the initial interviews to identify commonalities across the space science topics (and the differing levels of detail covered by each instructor), I designed a misconception survey to identify which concepts students continue to struggle with, even at the college level.

3.1.2.1 Naïve Physics

Previous work has explored the impact of students' conceptual representations of physics and space concepts, especially for more abstract concepts (Reiner et al., 2000). Yair, Schur, and Mintz (2003) discuss other examples which are famous in the scientific misconception community, including the inability of Harvard graduates to explain the cause of the seasons. They propose using scientific visualization tools and VR combined with a thinking journey, where the student can observe phenomena occurring and change their common misconceptions (Yair et al., 2003).

The Astronomy Diagnostic Test (ADT) is an assessment developed to assessment the baseline knowledge students in introductory college classes may have (The Collaboration for Astronomy Education Research (CAER), 1999). It has been found to be a reliable and valid assessment, though it focuses mostly on high-level questions related to the seasons, orbits, gravity, and more complex ideas like Kepler's Laws (Deming, 2002; Zeilik & Morris, 2003). Due to the specific nature of the questions for the ADT compared to the broader information discussed by the astronomy instructors, I designed a separate survey to measure more general astronomy knowledge from students.

3.1.2.2 Astronomy Study

A 28-question exploratory survey (available in Appendix B) was designed to identify which content students remember or struggle to understand, even after covering a basic introduction to astronomy throughout their K-12 (and possibly college) schooling. The survey covered three main areas, which generally came from the interviews completed with astronomy instructors, including:

1. Basic knowledge of the planets (# of planets, their order, size & scale)

2. Planetary details (including their unique features)

3. Conceptual questions (about gravity, seasons, and moon phases)

Two additional components were: 1) a drawing representing the solar system to scale and why they chose that representation and 2) their confidence levels for each question.

Each of these questions was exploring areas which the teachers in Study 1 identified as being common misconceptions or areas where students lack content knowledge. The final question (asking the participant to draw the solar system to scale) meant to provide insight to the participant's mental model which they may have been picturing while answering the other questions, and to see if there was a relationship between incorrect questions and the model they chose.

Sixty-nine undergraduate students ages 18-25 were recruited through the Psychology Participant Pool at Georgia Tech and completed the survey. A post-survey demographics collected information about their majors, minors, the last class they took which covered astronomy, and any informal learning activities they participate in. Participants had one hour at most to complete all of the questions and were instructed to leave no questions blank (i.e., give their best guess for each one). This questionnaire was administered on paper, and they could not return to previous questions after completing a page (some content in later questions could impact their responses on earlier ones).

Each survey was scored by allotting one point per correct answer, and the grading key was created before scoring any questionnaires. Since each section had different numbers of questions, percentages of correct answers were compared. The mean percent correct for the general solar system questions was 52.72, planetary details were 57.68, and conceptual questions were 63.04. There was a moderate correlation between scores for all three question types: solar system and planetary $r(67) = 0.47, p < .05$, solar system and concept $r(67) = 0.41, p < .05$, and planetary and concepts $r(67) = 0.42, p < .05$.

Misconceptions parallel to those identified by the instructors were found through the participant's responses to the questionnaire. For example, many participants thought the phases of the moon were caused by Earth partially blocking the Sun's light, instead of a deeper understanding of the relative positions from the sun and moon. For the last question (drawing the solar system to scale), over 95% of participants drew the planets in a linear order moving out from the sun, and many of them struggled to correctly reflect the differences between planet size (e.g., Earth vs. Jupiter vs. the Sun) and only one student had the correct distances scaled between them. Understanding what mental model students rely on when they are thinking about the solar system can help to design educational materials which might cause them to metacognitively reflect on their knowledge and engage with the content more deeply.

While many participants got at least half of the answers correct on the questionnaire, many of them would not receive a passing grade on a real test about these concepts. More worrying is the carry-over of misconceptions (e.g., about the seasons or moon phases), which are continually addressed at all levels of education; even students who may have covered this material multiple times could not accurately or logically answer those questions. Additional materials, including other modalities which may not rely on visual representations, could help address these misconceptions and may be able to help students explore and learn in a meaningful way.

3.1.3 Solar System Sonification Design

After collecting background information about this problem space through interviews and the misconception survey, we decided on the information mappings to highlight in the solar system sonification, including details relating to size and scale broadly, and individual planetary features.

3.1.3.1 Background (space sonification)

Sonifications have been used in a variety of applied contexts, including data exploration and education (Walker & Nees, 2011). There is some precedence for using auditory displays to analyze space and astronomy-related data, though many have not included formal evaluations. xSonify was one tool designed to create a sonification environment for analyzing space data (Candey et al., 2006; Diaz-Merced et al., 2011). Landi et al. (2011) completed an analysis of solar wind data through the audification (direct mapping of a dataset to sounds) of solar rotation data in order to explore carbon ionization. Lunn and Hunt (2011) have used sonification and audification as a way to analyze data sets from the Search for Extra-Terrestrial Intelligence (SETI) and the Cosmic Microwave Background Radiation.

Many applications of auditory display to astronomy have been completed mostly for outreach purposes, including Harger and Hyde's work to broadcast sounds from radio telescopes over the radio and internet (Ballesteros & Luque Serrano, 2008; Harger & Hyde, 2004). More recently, some have created solar system-centric sonifications. Ballora (2014) designed a musically-composed sonification for an outreach film presentation for the Smithsonian Air & Space Museum, while Quinton, McGregor and Benyon (2016) developed a solar system model after interviewing a planetarium expert. In this work, they identified seven properties to include in their model including length of day, gravity, and orbital period (year). Their evaluation included interviews with 12 users who were asked to provide their interpretation of the model's mappings without any scaffolding or contextualization.

3.1.3.2 Model Description

Work by Tomlinson et al. describes the designed solar system sonification more in-depth, though an overview about its design will be explained here (2017). The solar system sonification included two views: a Solar System Perspective (SSP) and a Planetary Perspective (PP). The SSP focused on presenting a baseline set of information for the audience. It also included comparisons of the size and scale for each object, and included each planet's mass, length of year, length of day, and distance from the sun. The PP provided information local to each planet (number of moons, rings, gravity, and mean temperature) and presented comparisons between these details. Data for each of the sonification mappings came from NASA's Planetary Fact Sheet (Williams, 2015). Each portion of the sonification was carefully designed to scaffold the audio-only comparisons, and to highlight details & concepts participants might not have previously known. Details were introduced in short chunks and were grouped by topic; some pairs were played together to support easier comparison between features.

3.1.3.3 Background (mapping sonification literature)

The design of each sonification mapping was informed by previous sonification research and our own design preferences. The SSP included mass, length of day, length of year, and distance from the sun (Tomlinson et al., 2017). Mass for each planet was created from brown noise which using a resonant filter whose center frequency was scaled proportionally to the mass, based on previous polarity mapping research (Walker, 2002a). Length of day was represented through a modulated amplitude envelope where the volume would move between zero and full amplitude. Each day would start from zero, increase as it moved toward sunrise, and decrease in volume till sunset. Here the goal was to scale 24 hours into one second, following a tempo mapping (Flowers, 2005). This created a perceivable pattern for all planets except for Mercury and Venus, whose days are very long compared to Earth's (about 58 and 116 respectively).

Length of year took advantage of the spatial audio setup available at the Fernbank Planetarium; since they had a quadraphonic (4-speaker) setup, it was possible to move each planet from speaker to speaker using Vector-Based Amplitude Panning. For this mapping, the length of year was represented by the speed at which each planet moved around the listener. Two different reference planets were used, due to the scale differences (Mercury for the inner planets and Jupiter for the outer).

The final representation for the SSP was the distance from the sun. Tomlinson et al. created a spacecraft sound containing additional sound effects for passing planets and asteroids (2017). An increased playback rate (which increased the pitch of the ship as it traveled) paralleled the speed of the ship as it moved from one planet to another. The base sound was used for the first four planets, then it ‘accelerated’ (through pitch shift) five times as quickly to reach Jupiter and Saturn, then ten times as quickly to reach the outer planets. This distance was logarithmically scaled to fit the large scale for the solar system (Tomlinson et al., 2017).

The PP included five different details: number of moons, number of rings, mean temperature range, gravitational strength, and type of planet (Tomlinson et al., 2017). Mapping strategies for the number of moons came from more traditional mapping strategies for physical qualities, where the number of tones represented the number of moons (one sound per moon); each pitch was randomly sampled from a range higher than the pitch for Mercury, but did not do a direct mapping since we were trying to emphasize the total number of moons, not make direct comparisons between each moon’s size (Dubus & Bresin, 2013; Tomlinson et al., 2017; Walker, 2002b).

The number of rings for each planet was represented through tones, with each additional tone representing another ring; in order to represent the idea of envelopment which you might get from a visual representation of rings, we used equal amplitude (loudness) through all four speakers to encase the audience in them. More rings resulted in an overall louder representation; Saturn’s rings were the loudest since it has the most layers (Tomlinson et al., 2017). The mapping for mean temperature range followed typical mappings from previous work, where low pitch represents colder temperatures and high pitch represents warmer ones (Dubus & Bresin, 2013; Flowers, 2005; Kramer, 1994; Walker, 2002b). In order to account for the very cold ice giants (Uranus and Neptune) and the extreme heat on Venus, each planet’s temperature was normalized and given a minimum frequency of 200 Hz (Tomlinson et al., 2017).

Gravitational strength was meant to be a complementary mapping to mass (since the two concepts are related); this time we used a model of a physical bouncing ball, where the pitch was proportional to the size of the planet, and the bounce rate was directly determined by gravity (Tomlinson et al., 2017). As many people already have mental models about this phenomenon, through experiencing it on Earth, we hoped to leverage this mapping as another way to represent size and scale. The mapping for type of planet (composition) tried similar things. It was meant to both be unique to each planet (each base sound started from the initial mass mapping), but also have a noticeable pattern between terrestrial and gas planets. The sound for the gas giants (e.g., Jupiter) was more diffused, and had more echo, while the sound for the terrestrial planets (e.g., Mars) had less echo and a higher density (Tomlinson et al., 2017).

In some instances throughout the show, a few of the sonifications would be played simultaneously, taking advantage of Auditory Scene Analysis (and its ability to help us parse multiple data streams), to scaffold comparisons directly between two plants, along one data dimension (Bregman, 1993; Schuett et al., 2014). A detailed outline of the solar system sonification script is available in Appendix C.

3.1.3.4 Methods

Evaluation of the solar system sonification took place at the local Fernbank Science Center, in Atlanta, Georgia. The show was free to attend and was advertised both at Fernbank and at Georgia Tech's campus. While there was the possibility of some recruitment bias, since the show happened during a typical showtime at Fernbank, the audience was representative of a typical attendee. As attendees entered, they were asked if they would like to participate in the survey during the show, to provide feedback. Attendees listened to the first part of the show (the SSP), then completed the survey, and then repeated this process for the second half (PP). For more details about the solar system sonification, see Tomlinson et al. (2017).

Forty people (ages 11 to 63) completed the survey during the show to provide feedback, and about half of them (19) were students. Attendees provided high-level feedback about the SSP and PP, and then answered open-ended questions about their most and least favorite parts of the show. After each section of the show (for the SSP and PP), attendees provided responses to two groups of Likert-type questions addressing aesthetics (two questions) and usefulness (three questions). The responses to each question were given six-point anchors to encourage the attendee to select either a positive or negative rating for the user experience, instead of choosing an easier neutral option. The Likert-type items provided a way for the audience to rate aesthetics and ease of use (comprehensibility) of the SSP and the PP, factors which may influence their overall listening experience and the amount of engagement they might feel listening to the show. A low comprehensibility or aesthetic score may mean more disengagement, disinterest, or confusion from the audience, and a worse overall listening and learning experience; high scores on those questions represents a better experience (and possibly better engagement in the learning experience).

The open-ended questions were designed to solicit audience feedback on other factors which may influence their engagement and listening experience during the show, including providing open-ended feedback about their affective responses to the auditory display. Feedback on their affective state during the show can provide insight to an audience member’s emotional engagement with the material. See Appendix C for a complete list of questions, including demographics, that show attendees answered.

3.1.3.5 Results

A short summary of the results will be given here, with the complete review available for more details (Tomlinson et al., 2017). Through the survey questions, the audience rated the overall aesthetics of the sounds as particularly high, with the ratings for both the SSP and PP being at least 4.7 out of 6. Audience ratings for usefulness in both the SSP and PP were high, though the PP had slightly lower ratings; this may be due to the more complex comparisons between each planet taking place in the PP. High ratings for aesthetics and ease of use provides insight to their level of engagement with the materials, as those factors can influence enjoyment, interest in the material, and concentration on the show.

At the end of the entire show, the audience completed free response questions asking for their favorite or least favorite sounds, if they learned anything new during the show, and if they had any affective, or emotional reaction, while listening to the show. Positive affective experiences relate directly with emotional engagement and could also affect learning. Many audience members enjoyed the sonification mappings representing the gas giants and any section of the show which compared between two planets (Tomlinson et al., 2017). They also reported a better understanding about the scale and relationships between the two planets (which was represented during the PP), and at least half mentioned learning new information about our solar system. Some attendees provided specific examples of how the PP supported comparisons between two planets, especially for weather and atmosphere (Tomlinson et al., 2017).

3.1.3.6 Discussion and Reflections on the Design

Overall, everyone in the audience really enjoyed their experience listening to the solar system sonification. Many attendees mentioned having more trouble remembering details from the PP, probably due to the amount of details (up to six) for each planet. The PP was also reported to be harder to understand than the SSP, which may be due to the amount of comparisons between planets during each section. Allowing someone to navigate at their own pace through these representations, or to group by detail (e.g., mean temperature) instead of by planet may help support comparisons more easily. One sound mentioned by multiple attendees as being unpleasing aesthetically was the sound for the moons; while higher pitched tones were used to represent them (to fit with the pitch-to-mass mapping used for the planets), they were still quite high and could be a bit grating. Using a filter to reduce the pitch for this mapping would make it easier for someone to listen to those sounds.

3.1.3.7 Conclusion

Even though planetariums typically rely on a presentation of combined visuals and descriptive audio, the sonified planetarium experience was a successful deployment for enjoyment and learning. These results imply that people can enjoy and engage with alternative experiences in these environments, learn subject-specific details through sonifications, and have positive affective experiences listening to an auditory display (Tomlinson et al., 2017).

3.2 Study 2: Listening Activity Evaluation

3.2.1 Purpose

While the planetarium deployment was a successful first phase of evaluation for the solar system sonification, it did not attempt to measure (in a concrete manner) learning or display interpretation. This second study was an evaluation of the overall design of the solar system sonification, including measuring the ease of use and aesthetics of the display. It included a more structured listening task than the original planetarium deployment (to measure accuracy of display interpretation), a pre- and post-test to measure learning, and other questions to measure engagement, including standard measures of user experience.

3.2.1.1 Scale development

While many current UX scales can provide a general comparison between the usability rating of multiple systems, people usually have a low level of familiarity with auditory displays, which can cause varying levels of interpretation for the statements within the scale. I developed a scale composed of 11 statements to elicit feedback on data mapping interpretation and aesthetics: five items were inspired from work done on interpretation of peripheral displays (Matthews et al., 2007), and the last six items were for getting feedback on meaning, enjoyment, and comprehension (Tomlinson, Noah, et al., 2018). An auditory user experience scale could help evaluate someone's engagement with the auditory display.

3.2.1.2 Learning evaluation

A pre-test was used as a way to get a measurement for how much baseline knowledge each participant had about astronomy before listening through the solar system sonification. Ten questions, covering varying levels of detail about the solar system, were given to each participant through Qualtrics. Questions probed for different levels of detail including ones about size and scale (e.g., order of the planets from the Sun out, ranking them from smallest to largest) and also ones about defining characteristics for each planet (e.g., selecting distinctive features about Venus).

After completing the listening activity questions (for the SSP and PP), each participant completed a post-test using the same questions from the pre-test. Self-reported confidence scores were asked for each question, to prompt each participant to think about their current level of knowledge for each question before the learning activity and afterward. Differences between the pre- and post-test scores were used to measure learning outcomes from listening to the solar system auditory display, with more potential for learning occurring when someone is engaged at a cognitive and/or emotional level.

3.2.2 Methods

3.2.2.1 Participants

A total of 52 participants (20 females and 32 males) with an average age of 20.1 (SD 1.7) from a large research university in the United States took part in the study. All participants were required to have normal or corrected-to-normal vision and hearing. Participants reported their majors, the last class they had astronomy in, their typical attendance of informal learning activities, and online/print media they follow related to astronomy. Each of these factors may have resulted in a large amount of prior knowledge which could potentially skew the data resulting in high pre-and post-test scores.

3.2.2.2 Materials

Audio stimuli were presented using Sony MDR-7506 Studio headphones. Participant responses were collected in a computer lab, with each student working at their own pace. A previously-designed recording of the solar system sonification was used as the referent auditory display for all trials (Tomlinson et al., 2017). Participants in this study listened through the two-part recording containing the SSP and PP.

3.2.2.3 Study Design

Participants completed the 10-question pre-test (and the confidence ratings for each of those questions), then moved on to the listening activity. They listened through the first half of the solar system sonification (the SSP, about 16 minutes long) and then answered specific questions relating to the information covered in that half of the display. After answering topic-specific questions, each participant responded to the user experience questions, including the UMUX, the audio UX scale (BUZZ), and open-ended questions asking about their overall likes and dislikes for the displays and mappings in that section. Then they listened through the rest of the display (the PP, about 11 minutes long), and completed similar topic-specific questions and UX questions. The questions in the PP section of the listening activity were more complex, and a few required some more in-depth interpretation of the displays compared to the simpler interpretations from the SSP. Then each participant completed the post-test and an exit survey about their informal science experiences, astronomy background, and general demographics. Each session lasted anywhere from 30 minutes to one hour, depending on their pace. All questions asked during the study are available in Appendix D.

3.2.3 Analyses and Results

Each response for the pre- and post-test, as well as the listening activity question responses, were recorded through Qualtrics. The pre-post scores were analyzed through a paired samples t-test, to explore whether or not there was a significant change in the overall number of correctly answered questions after completing the listening activity. Overall confidence ratings for the responses on the pre- and post-tests were analyzed through a paired samples t-test, too.

There was a statistically significant difference between the total scores on the pre-test ($M = 13.25, SD = 3.75$) and the post-test ($M = 16.14$, $SD = 3.27$), $t(51) = 6.544$, $p < .001$. From this, we can infer that there was learning from the content in the solar system sonification, as scores improved. There was also a statistically significant difference between the overall pre-test confidence scores ($M = 32.75$, $SD = 6.86$) and the post-test confidence scores ($M = 40.81$, $SD = 4.52$), $t(51) = 14.7$, $p < .001$. From this, we can infer that there was an impact of the solar system sonification on confidence levels.

Analyses for the user experience scales (UMUX and BUZZ) provided insight for each participant's engagement with the materials. Before analysis of the user experience scores for the audio UX scale (BUZZ), each negatively worded item was converted to the same scale as the positively worded ones, by subtracting each value from eight. Summing the total score for each item provides a total score out of 77. More details about using this scale are available in Tomlinson et al. (2018).

A Principal Factor Analysis (PFA) using Varimax rotation with Kaiser Normalization was completed for both the SSP and the PP. Two factors were found for the

SSP, one factor contained items related to enjoyment and appeal (aesthetics) while the second factor contained items related to ease of use (Tomlinson, Noah, et al., 2018). The PP PFA resulted in three factors and broke up the items relating to ease of use in the SSP into two groups: one group related to ease of use while the second included items related specifically to understanding (e.g., 6. 'It was easy to match these sounds to their meanings.') (Tomlinson, Noah, et al., 2018).

A correlation of the overall BUZZ score for the SSP and PP was completed with the UMUX scores for those sections, as a way to evaluate the validity of this new scale. For the SSP, the BUZZ score correlation with UMUX was $r(50) = 0.68, p < .001$; for the PP, the BUZZ correlation with the UMUX was $r(50) = 0.74, p < .001$ (Tomlinson, Noah, et al., 2018). In addition to evaluating the validity through its correlation with UMUX, a reliability statistic, Cronbach's alpha, was calculated for each of the factors and the entire set of statements for each perspective (for reliability ratings, see Table 1).

Table 1 - Reliability summary table for the overall score and the SSP and PP individual factors.

	Factor	Items	Alpha
Solar System Perspective (SSP)	Enjoyment and Appeal	1 - 3, 8, 9	0.88
	Ease of Use	4 - 7, 10, 11	0.85
	Overall	1 - 11	0.88
Planetary Perspective (PP)	Enjoyment and Appeal	2, 3, 8, 9	0.91
	Ease of Use	1, 4, 7, 10, 11	0.86
	Understanding	5, 6	0.69
	Overall	1 - 11	0.83

3.2.4 Discussion

There was a significant improvement between the pre-test and the post-test scores, meaning that there was learning from the listening activity. There was also a significant improvement in overall confidence scores for the post-test compared to the pre-test, which means that listening to the solar system sonification impacted not only overall accuracy on the tests but resulted in a participant having higher confidence in their own answers as well. Self-confidence (efficacy) in learning can have a large impact on cognitive engagement, and high post-test and post-test confidence scores provide support for the auditory display's ability to engage participants in a meaningful way.

The outcome from the factor analysis found that the SSP had two factors and the PP had three. One reason for this may be that there were major differences between the task difficulty in the second half (the planetary perspective), or the types of data represented were very different from data in the first half (Tomlinson, Noah, et al., 2018). Instead of including only information about size and scale (the way the SSP did), the PP included many more details about each planet, and were presented grouped by planet. Allowing for the separation and comparison of one variable to another (e.g., comparing mean surface temperature on Venus with Earth, instead of moving between planets and giving an overview of all of the details for one planet at a time) may mitigate potential difficulties. The development of a standard measure for audio user experience provides a way to evaluate cognitive and emotional engagement, through probing enjoyment and appeal, as well as ease of use.

3.2.5 Conclusion

After the original deployment of the solar system sonification at the Fernbank Science Center, it was important to evaluate whether or not people could actually learn while listening to an audio experience like this, instead of evaluating only their listening experience. This lab study provides the first evaluation of the ability for an auditory display to support significant learning outcomes for astronomy content, and explores engagement through self-efficacy, emotional engagement, and cognitive engagement.

Additionally, this was preliminary work to develop a scale for evaluating the usability of an auditory display. The overall reliability measures were pretty high (0.83 and 0.88); further testing with the BUZZ audio UX scale would provide additional information about how well it is measuring the understanding the listening experience for a user. Follow-up studies correlating the BUZZ score with the SUS should also be done, to provide another validation measure in addition to the UMUX. Other comparisons between UMUX and the SUS have been completed (Finstad, 2010), and asking both the SUS and UMUX would provide two opportunities for comparison against validated scales, in addition to a way to measure the internal validity of how well participants are interpreting the responses on the SUS and the UMUX (i.e., since they have been found to be highly correlated, if the SUS and UMUX scores for a study do not follow that, there are differences in interpretation happening between users for each of the items).

While this scale was validated on a small scale (with 52 participants) further research should be done validating it. Additional testing should be done to see if the scale is internally valid (i.e., if this study were replicated, do the SSP and PP have the same factors occur during a factor analysis?). When studying comprehension of complex auditory displays, Schuett and Walker (2013) described other possible evaluations including the need to measure ratings of workload (e.g., through NASA TLX (Hart & Staveland, 1988)), situation awareness, and measures for accuracy and latency of response time. Measuring subjective workload, like NASA TLX may provide another useful benchmark for understanding and interpreting the BUZZ audio UX scores.

CHAPTER 4. STUDY 3

4.1 Participants

Middle school students ages 12 to 14 were the target group to participate in this study; however, as the recruitment for this age was conducted through smaller groups of students from local organizations with more diverse age ranges, the study was opened to all learners aged 12 to 17. Participants were recruited through word of mouth, email, posting to a local Tech404 slack channel in the Atlanta area, and through the Center for Education Integrating Science, Mathematics, and Computing's (CEISMC) First Lego League, both in person and over email.

Seventeen students participated in the study. Fourteen of the study sessions took place in the Psychology Building on Georgia Tech's campus, and another three participated at a local private school in the Greater Atlanta area. The average participant age was 13 (SD = 1.029), and 10 participants self-identified as male and seven self-identified as female. Their average reported most recent space-related science learning experience was in 5th grade (from 11 participants); six participants could not recall their most recent space-related learning experience.

4.2 Materials

4.2.1 Universe Sandbox Designs

A deployable version of Universe Sandbox[1] (Giant Army, 2015), a space interactive simulation, was used by every participant in this study. Four different conditions were compared using a between-subjects design. The four conditions differed by tool type (PC vs. VR interactive simulation) and number of modalities (visual only vs. visual + audio). The multimodal conditions used sonifications to represent different data mappings for the solar system.

Source code was provided by Giant Army along with permission to adapt the code for both the PC and the VR conditions. All code was modified using Unity[2] and SteamVR[3], and generated sound clips were integrated using the built-in 3D audio functionality supported by Unity. The installed version used for this study was modified to start with the Solar System pre-loaded and removed much of the editing & creation functionality to constrain the exploration environment for learners. No large changes were made to system functionality, usability, or general user interface layout.

The PC conditions were run on a Dell laptop, while the VR conditions were run on a desktop (for participants at Georgia Tech) and an MSI gaming laptop (for participants at the private school) using a Vive Pro VR headset[4].

---

[1] http://universesandbox.com/
[2] https://unity.com/
[3] https://store.steampowered.com/app/250820/SteamVR/
[4] https://www.vive.com/us/product/

4.2.1.1 PC Versions

The PC – audio and no-audio versions used the same visual display, and both opened to a view of the solar system with a modified set of controls at the bottom (Figure 1). When a planet is selected with a single left-click, a small overview panel opens (Figure 2). Clicking inside of that small panel opens the larger detailed panel containing different categories of information in the overview, motion, composition, and temperature tabs (Figure 3). Double-clicking would set the selected object as the new central camera focus, and learners could play/pause the time, change the time scale, and zoom or rotate the camera view to explore.

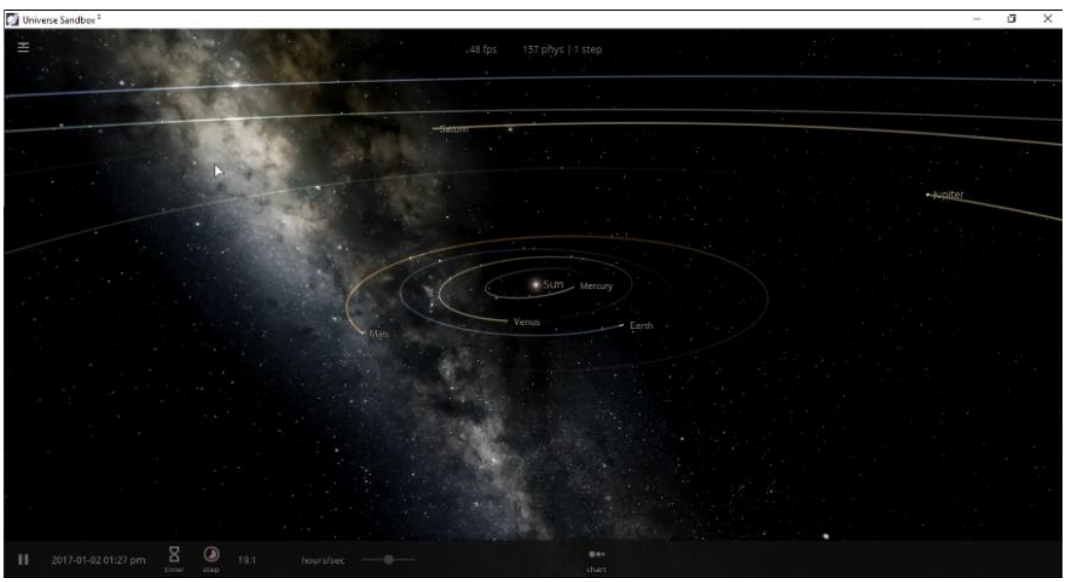

Figure 1. Initial PC view of the solar system when started.

Figure 2. Small overview panel for Mercury, containing a summary of mass, diameter, density, temperature, and velocity.

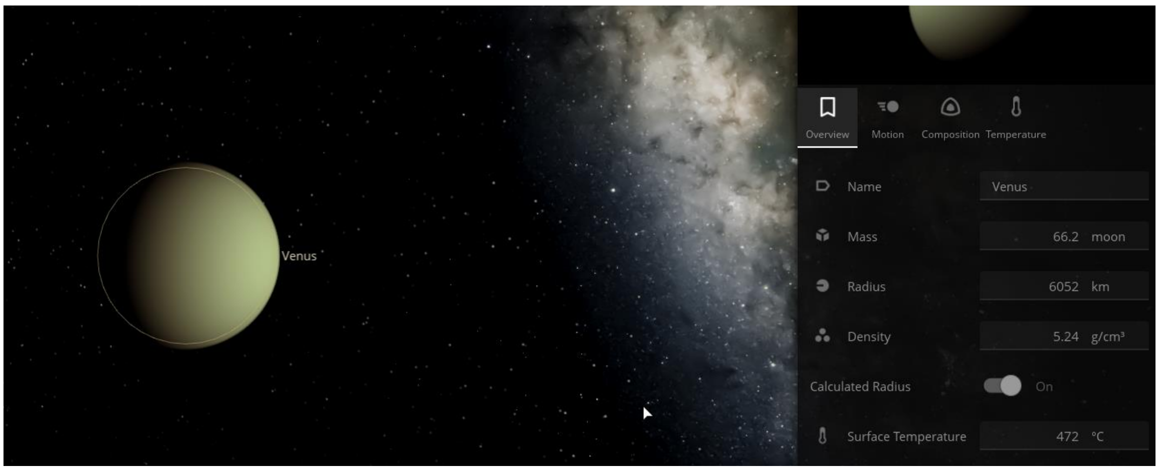

Figure 3. Detailed overview panel for Venus shows four tabs: Overview, Motion, Composition, and Temperature, with numerical values for mass, density, and surface temperature showing.

The audio version of the simulation used the exact same visual representations but included additional information through 6 types of sounds. The mass sounds play spatialized around the participant, at the position of each planet, based on the within-simulation camera location. The volume of all planets drops off as the participant zooms in to one planet, until only the mass for that focused planet is audible. As they zoom out, the other planetary mass sounds fade back in. The sound representing the presence of rings plays when a planet is double-click focused (if there are no rings, no sound plays). Each tab of the large planetary details panel has a sound associated: the Overview Tab played gravitational strength, the Motion Tab played length of day, the composition Tab played density/composition, and the temperature Tab played mean surface temperature. Each sound played on a loop, allowing the participant to hear them as many times as was useful. Closing the overview panel stopped the sounds.

4.2.1.2 VR Versions

The VR – audio and no-audio versions also used the same visual display, and also opened to a view of the solar system (Figure 4). Participants used one of the two controllers during the duration of the study. They were able to switch between three movement modes: Teleport (achieved by pointing at a target, waiting for the label to pop-up, then pulling the trigger), Fly (achieved by pointing the controller pointer in the desired direction and holding down the trigger), and Grip Button movement (achieved by pushing in the side grip buttons and moving the controller around, similar to panning). Changing between movement modes Teleport and Fly, changing the time scale, and resetting the simulation were available through the trackpad (Figure 5). After teleporting to a planet, hovering the pointer over a planet would show a similar small overview panel with details (Figure 6).

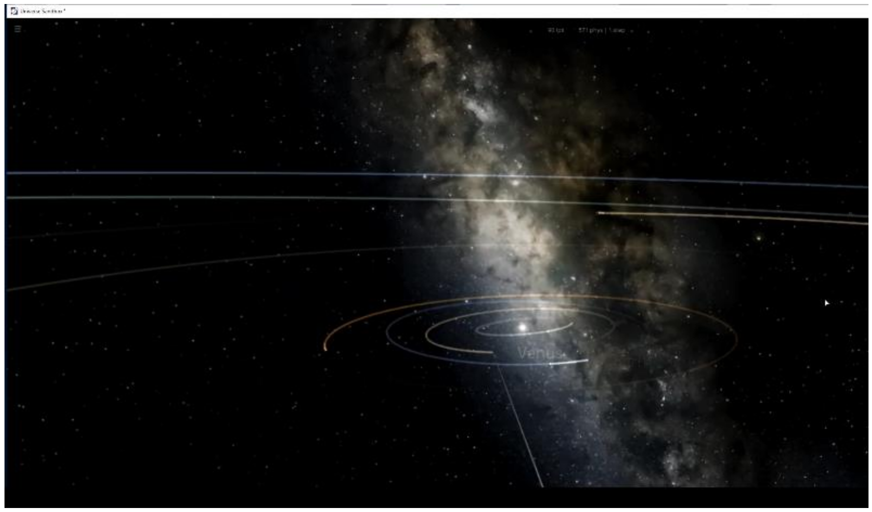

Figure 4. Initial VR view of the solar system when started.

Figure 5. Vive Pro Controller[5].

[5] https://www.vive.com/us/accessory/controller2018/

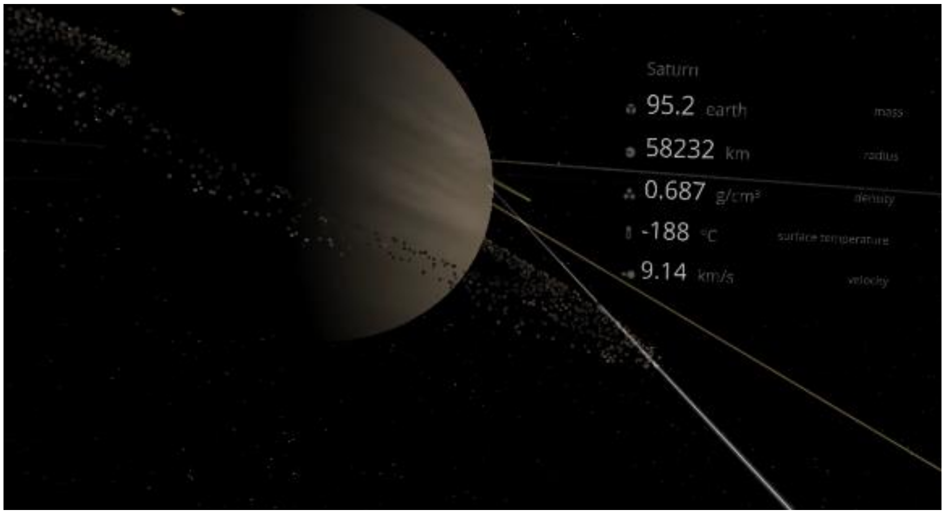

Figure 6. Overview panel for Saturn shows on pointer hover.

The audio version of the VR simulation used the exact same visual representations but included the same 6 types of sounds as the PC – audio version. The mass sounds play spatialized around the participant, at the position of each planet. The volume of those planets drops off as the participant teleports to one planet, until only the mass for that focused planet is audible. If they fly or use the grip button to move, or reset the sim, then the other planetary mass sounds fade back in. Teleporting to a planet triggers the ring sound (if there are no rings, no sound plays). Hovering over a planet to show the overview panel reveals an additional button on the trackpad ("Play Sound"). With the panel showing, pressing the button will play each sound in order, and color the associated numerical value teal to provide feedback: mass played gravitational strength, density played density/composition, mean surface temperature, and velocity played length of day (Figure 7). Sounds repeat until the selector is moved to the next detail and stop when the pointer is moved away from the planet.

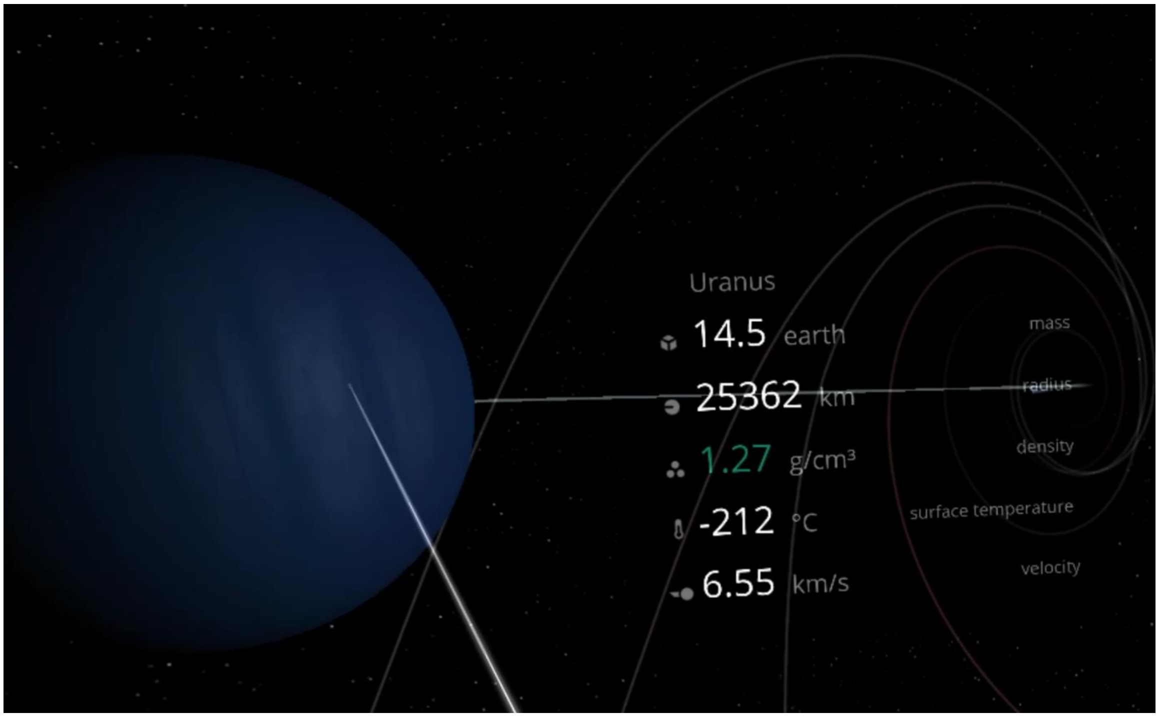

Figure 7. Overview panel for Uranus shows on pointer hover, playing the density sound.

4.2.2 Pre-Activity Surveys

A variety of pre-activity questionnaires were asked before using the Universe Sandbox. These included the Adapted Technology Experience Profile (Liles, 2017), the modified Attitudes Toward Science Inventory (mATSI) (Weinburgh & Steele, 2000), the Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988), the pre-activity solar system questionnaire, and for the two VR conditions, the Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993). Pre-activity materials are available in Appendix E.

4.2.3 *Activity*

Screen capture and audio was recorded from each participant session, in both the PC and VR conditions. Participants completed a semi-structured activity where they explored the Universe Sandbox. During this activity, they answered questions about the solar system (e.g., “Explore and discover what the two coldest planets are”). These questions were adapted from the listening activity study. This activity was designed to be inquiry-based (as these can support deeper cognitive engagement), and participants were scaffolded through activity questions which relied on contrasting cases, and asked the participants to discover what the hottest and coldest planets were; others prompted them to compare length of days and find a planet with one shorter or longer than Earth’s. These questions encouraged participants to explore and make deeper comparisons, instead of merely reporting a single value each time as a sufficient answer. At the end of the activity, they were given a few minutes to explore and play with the system however they wanted. Activity materials are available in Appendix F.

4.2.4 *Post-Activity Surveys*

A variety of post-activity questionnaires were asked after using the Universe Sandbox. These included a post-activity solar system questionnaire, PANAS, the Slater-Usoh-Steed (SUS) presence scale (Slater et al., 1994), UMUX (Finstad, 2010), the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991), the Science Activity Questionnaire (SAQ) (Meece et al., 1988), and for students in the multimodal conditions, the BUZZ audio user experience scale (Tomlinson, Noah, et al., 2018). Participants also answered six open-ended reflective questions about the experience and demographics. Post-activity materials are available in Appendix G.

4.2.5 Follow-up Interview

About one week after the study, participants completed short (10-minute) optional follow-up interviews. If the participant was interested in completing the follow-up, the interview was arranged before the participant left the study session. The interview was semi-structured, and 11 questions were asked, including “Can you describe what the experience with the simulation was like” and “Can you describe a detail or two about one of the planets?” Questions gradually moved from more general to more specific to prompt participants to share early-on the most salient, memorable experiences before biasing their responses by asking for directed answers. Follow-up interview materials are available in Appendix H.

4.3 Procedure

On-campus participants came to the Psychology building at Georgia Tech. A small stipend was available to offset the cost of transit to campus, and parking costs were covered for each participant. After receiving parental consent and student assent, each participant was randomly assigned to one of the four conditions. They completed the four (or for VR, five) pre-activity questionnaires in the following order: Adapted Technology Experience Profile, mATSI, PANAS, pre-activity solar system questionnaire, and if applicable, the SSQ.

The Adapted Technology Experience Profile asked them to report which technologies they are familiar with, across a variety of contexts including home and school (see Appendix E for these items). Then they completed the mATSI to get a baseline for their current attitudes about science. Next, each participant completed PANAS survey to forecast their own affect of using their particular version of the Universe Sandbox; the prompt text was modified to include information about the simulation version they used. Finally, each student completed the pre-activity solar system questionnaire for a baseline of their astronomy subject-area knowledge.

After the pre-activity questionnaires were completed, participants were trained on how to use the system; at this point, the sounds were muted for the audio-enhanced conditions. Then, participants in the audio conditions were also given a short introduction to the sound mappings outside of the simulation context. Following that, they completed a short tutorial with the sound un-muted. Before the activity started, participants could ask any additional questions about controls, the task, or the sound mappings. At this point, students in the VR conditions completed the SSQ, and were asked if they had any problems continuing with the study.

Each student participated in a 20-minute explorative learning activity using their assigned version of the Universe Sandbox. The semi-structured activity question prompts were used to encourage the students to explore a variety of planets, details, and gave them time to become familiar with the software.

Upon completion of the learning activity, participants answered the post-activity astronomy questionnaire. Then a follow-up PANAS questionnaire was given to measure each student's current affect after using Universe Sandbox, and the SUS questionnaire was given to measure immersion and presence they experienced during the activity. Participants then reported overall user experience using the UMUX questionnaire. They completed the MSLQ and SAQ surveys to report motivation, self-efficacy, and engagement. Participants in the two audio conditions then answered BUZZ to report on auditory user experience. Finally, each student answered six open-ended questions reflecting on their experience, and completed a demographics questionnaire. Each session took approximately 1 hour in total.

Data was also collected off-campus, at a Greater Atlanta Area private school. Due to additional time constraints of students (only 15-20 minutes total were allowed), the instructions were slightly condensed for the two audio conditions (as those had the longest set of instruction). Instead of having a short exposure to the sounds outside of the simulation, they were introduced to them during the initial simulation introduction and orientation to controls. Additionally, as the pre- and post-surveys took the majority of the time, participants were given the pre-activity survey ahead of time & brought it in to the study session. After the session, they were given an electronic version of the survey to complete. Follow-up interviews were still scheduled at the end of the in-person session.

4.4 Pilot Testing

Pilot testing was completed with eight Georgia Tech students (both undergraduate and graduate) as a way to narrow down to the final set of surveys (and time) for the study to help reduce the chance of fatigue. It was also used to evaluate the subjective workload between the PC and VR conditions, as it was unknown whether or not the headset would cause discomfort over a longer period of use. Average physical workload was measured through NASA-TLX (Hart & Staveland, 1988). The mean score for Physical Workload was 13.475 (*SD* = 20.192), and there we no large differences between groups, with one notable exception for one pilot participant in Group D (who gave the highest workload score: 51.7). Since overall Physical Workload was relatively small, the length of the activity (i.e., 15-20 minutes) seemed appropriate for the younger participants in the main research study.

Pilot testing was used to select the final set of surveys measuring engagement and motivation. For example, the SEI was given in the post-activity questionnaires, to measure self-reported motivation and engagement; however, participants thought the overly-general nature of the questions (e.g., "My teachers are there for me when I need them.") were not specific enough to science or the specific study activity, so they were dropped from the post-activity materials. Both the MSLQ and the SAQ were given to measure engagement and motivation with science activities. Pilot participants thought these questions fit better than the SEI, so these were both kept in for the final study. Participants also answered the mATSI before completing the simulation activity, and there were no concerns or confusion from participants when reviewing the question statements.

Participants answered questions from the original versions of the surveys and discussed any items which caused them confusion (e.g., due to wording, applicability, etc.). This feedback was used to make slight changes between the original scale wording, which focused on "classes" or "course materials," and the wording used in the final study, which focused on the "learning activity."

One final change was made to the study materials after piloting. Prior to pilot testing, the entire sonification set was added to the study. Participants reported dislike over the moon representation sound, as some moons were visible for each planet, unlike rings, which were only visible for Saturn. They preferred the rings sound, as it gave them additional information. The final version of the Universe Sandbox was updated to remove the moon sounds, and the rings were left in, as a direct detail comparison between the audio and no-audio simulation versions.

4.5 Research Design

Study 3 investigated three research questions, all focused on comparing experiential differences between the four conditions. These conditions are listed in Table 2. Each participant experienced one of the four conditions.

RQ1. Does a VR simulation or PC simulation support higher levels of emotional, intellectual, and physical engagement?

RQ2. Does an audio-enhanced simulation or a visual-only simulation support higher levels of emotional, intellectual, and physical engagement?

RQ3. What factors, such as technology experience, math and science anxiety, self-efficacy, and affect, influence a student's ability to interact comfortably with multimodal science tools?

Table 2 - The four study conditions.

	PC Simulation	VR Simulation
No Sound	Group A	Group C
With Sound	Group B	Group D

4.5.1 Hypotheses

It was expected that the VR conditions would have higher levels of engagement compared to the PC conditions, particularly in the physical and cognitive engagement scores. Additionally, based on the Multimedia Principle, it was expected that the multimodal (or "audio") conditions would have higher levels of engagement than the visual only ("no-audio") conditions.

The second research question explored whether or not the multimodal interactive simulations supported better learning opportunities than the more typical learning experiences (current non-audio simulations). It was hypothesized that the VR – audio group (Group D) would have the highest post-activity accuracy scores than all other conditions. It was also hypothesized that the two audio groups would have higher scores than their no-audio counterparts.

Finally, research question three was more exploratory and was meant to understand whether or not students may have different comfort levels or anxieties using newer technologies in the classroom. Understanding the effect of prior experience and student comfort levels with technology before completing a science activity can inform scaffolding of activity structure, instructions, and constraints for actual classroom use.

4.5.2 Analyses

Both quantitative and qualitative analyses were used to answer the research questions. Descriptive statistics were used to compare average scores between groups. Additionally, the Mann-Whitney U non-parametric test was used to evaluate differences between pre-, during, and post-activity surveys for the VR and PC groups, and the audio and no-audio groups. Qualitative analysis was completed through coding of the screen and audio recordings for each participant to compare differences between participants. Process, or action, coding (Charmaz, 2002; Corbin & Strauss, 2008) was used to complete the qualitative analysis, following the method presented by Saldaña (2013, pp. 96–100). Follow-up Chi-square evaluations to check for code frequency differences between the PC conditions, the VR conditions, and across all four groups.

To generate the codes, a set of four videos, one from each condition, were reviewed to observe actions and brainstorm the potential process codes. Observations and behaviors were recorded without tracking verbalization connotations. Then, these notes and actions were reviewed, and each item was summarized onto a sticky note in shortened form (2-5 words each). These notes were then categorized through affinity diagramming (Kokogawa et al., 2012), to create high-level groups and identify repeated/overlapping codes. Overlapping codes were considered and reduced into a single code, or made more specific, depending on the nature of the overlap. These codes were then organized into a set of seven categories. The categories were then reviewed, and checked for within-category connections, context, and structure. After that review, five categories remained: 1) Interacting with Planet Information and Details; 2) Using Movement or View Controls; 3)

Verbalizing Comments and Observations; 4) Playing or Replaying Sounds; and 5) Completing Non-Task Play (see Appendix J for a detailed overview of these codes).

Each code was then reviewed for usefulness (how much they detailed a behavior or action) and a description or set of "identification" rules. Some codes, such as "failing to zoom," were removed due to their potential for being coded inconsistently across conditions or participants. Codes were then labeled as being relevant for PC-only, VR-only, or both. Then all codes, their descriptions, and their high-level groups were reviewed with another researcher to verify their distinctiveness and categorical appropriateness of all codes. Videos were then coded using Atlas.ti[6]. The "zoomed" code for was removed from the VR conditions, as it was initially included in the lists for both PC and VR; however, it overlapped with the non-teleport movement, and would always be dual coded. Finally, some constraints were added, such as the length of time an action needed to take place before it could be coded as such. This was particularly relevant for the VR "hover" to open the information panel. If the participant was moving across a planet, the panel would open. Since that occurred as part of another movement (and not as a focused affixation on the particular planet for details), it was not marked with the related code. The general rule-of-thumb was that each action needed to occur for at least one second in order to be coded as such; this was often the minimum amount of time participants would spend viewing numerical details directing when completing intervals of viewing the planet for details and viewing the animations/other planets situated around their current focus.

[6] https://atlasti.com/

Each video was coded (through 3-6 passes, the total number of codes and often the length of video determined how many passes were used) and then reviewed before moving on to the next video. All videos were coded consecutively, starting with the first video in the PC – no-audio condition and ending with the last before moving on to the next group. Code counts were extracted from the software and compiled into excel after each video was completed. Average number of process codes were calculated for each group based on the extracted counts. The full overview of these counts is available in Appendix K.

CHAPTER 5. STUDY 3 RESULTS

The results from this study are split into the relevant data for each of the three research questions, although there is some overlap, particularly between RQ1 and RQ2. Two overview tables present all of the pre-activity survey and subscale scores (Table 3) and the post-activity survey and subscale scores (Table 4).

Table 3 - Pre-Activity Average Scores in order of participant completion: technology use (frequency profile and breadth), mATSI, PANAS, pre-activity, and SSQ. Each cell contains the mean and standard deviation.

		A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Frequency Profile	**General** (Out of 3)	1.01 (0.28)	1.07 (0.18)	1.44 (0.28)	1.47 (0.22)
	Game Use (Out of 3)	0.95 (0.17)	1.3 (0.36)	1.44 (0.41)	1.55 (0.26)
Breadth	**General** (Out of 36)	8.25 (2.49)	9 (2.55)	12.2 (1.47)	12 (0.71)
	Game Use (Out of 5)	2 (1.41)	2.5 (0.03)	3.4 (0.49)	3.25 (0.43)
mATSI	**Value of Science to Society**	18.75 (4.32)	20.5 (3.28)	21.2 (1.47)	22 (0.707)
	Anxiety Towards Science	9 (2.74)	8.75 (2.49)	9.8 (3.71)	7.5 (1.80)
	Self-Confidence in Science	19.75 (1.92)	21 (2.12)	21 (2.61)	20.75 (2.59)
	Desire to Do Science	24.25 (3.56)	26 (4.24)	24.2 (6.43)	29.75 (3.45)
PANAS	**Positive**	31 (3.94)	28.75 (7.79)	33.4 (6.02)	41.75 (3.49)

Table 3 - Continued

	Negative	13.5 (2.29)	15.5 (2.87)	13 (2.19)	13.5 (2.96)
Pre-Activity	**Correct**	3.5 (2.29)	3.75 (2.39)	6.4 (3.88)	6.5 (1.34)
	Extra Facts	4.75 (1.64)	6 (3)	4.4 (1.2)	5.25 (1.16)
SSQ		-----	-----	0.25 (0.433)	0.5 (0.5)

Table 4 - Post-Activity Average Scores in order of participant completion: post-activity solar system questionnaire, PANAS, presence SUS, MSLQ, SAQ, BUZZ. Each score is listed with the mean and standard deviation.

		A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Post-Activity	**Correct**	5.75 (2.86)	7.25 (4.21)	7.6 (2.58)	8.5 (2.86)
	Extra Facts	5.75 (1.48)	4.25 (0.43)	5 (1.1)	4.5 (2.05)
PANAS	**Positive**	31.8 (7.36)	30 (1.41)	28.4 (9.71)	43.5 (1.5)
	Negative	10.5 (0.5)	18.3 (8.26)	11.8 (2.23)	10.8 (1.3)
SUS		3.88 (2.22)	4.38 (0.25)	5.23 (1.15)	6.125 (0.22)
UMUX		17.75 (1.09)	18 (3.94)	19.6 (2.94)	21.75 (1.09)
MSLQ	**Intrinsic Goal**	3.88 (0.84)	5.31 (0.93)	4.55 (1.44)	5.63 (0.96)
	Extrinsic Goal	3.56 (1.72)	4.75 (0.90)	3.65 (1.64)	4.75 (0.88)
	Task Value	4.67 (1.26)	4.96 (0.82)	4.47 (1.4)	6.04 (0.59)
	Control Beliefs	4.69 (0.86)	4.94 (0.57)	4.25 (1.17)	5.63 (0.38)
	Self-Efficacy and Performance	4.97 (0.92)	4.59 (0.75)	4.55 (1.07)	5.09 (0.61)

Table 4 - Continued

	Test Anxiety	3.3 (1.27)	5.05 (0.64)	3.04(1.21)	3.1 (0.83)
SAQ	**Active Engagement**	2.03 (0.52)	2.49 (0.14)	2.24 (0.45)	2.75 (0.2)
	Superficial Engagement	2.4 (0.51)	1.4 (0.47)	1.24 (0.08)	1.2 (0.35)
	Task Mastery	1.78 (0.32)	3.16 (0.64)	3.13 (0.64)	3.91 (0.10)
BUZZ		-----	48 (9.25)	-----	55.7 (10.31)

5.1 Research Question 1

RQ1. Does a VR simulation or PC simulation support higher levels of emotional, intellectual, and physical engagement?

Multiple engagement factors were evaluated, including physical engagement (Presence SUS, process coding for tool use), emotional engagement (PANAS), cognitive engagement (SAQ, post-activity scores, during-activity scores, open-ended questions at the end), motivation (MSLQ), and longer-term engagement (follow-up interviews). The Mann-Whitney U test was used to evaluate differences between the VR and PC groups for each of the following factors: SUS scores, post-activity PANAS (both positive and negative) scores, SAQ subscale scores, post-activity accuracy scores, and the MSLQ subscale scores. Chi-square tests of independence were used to evaluate differences between the number of process codes assigned to each video to evaluate tool use.

5.1.1 Physical Engagement

The Presence SUS (Table 4 A) assessed each participant's sense of being immersed within a virtual environment (Slater et al., 1994). A high score represents a high sense of presence, with a maximum average of 7. There were group differences for presence in the virtual environment. The VR conditions had higher overall scores than the PC conditions. Group D had the highest presence score (6.13), while Group A had the lowest (3.88).

Table 4 A - Slater-Usoh-Steed Presence Questionnaire (Presence SUS) scores by group. Max value of 7; a higher score is better.

	A) PC - no-audio	**B) PC – audio**	**C) VR – no-audio**	**D) VR – audio**
SUS (SD)	3.88 (2.22)	4.38 (0.25)	5.23 (1.15)	6.125 (0.22)

A Mann-Whitney U test was used to evaluate group differences between the VR and PC groups. Participants in the VR group had a statistically significant higher presence score (Median = 6.0) compared to the PC group (Median = 4.42), $U = 14.5$, $p = .038$.

Physical engagement could also be described as the total number of actions from each participant when using the simulation (overview available in Table 5). Groups C and D had the largest number of interview sections assigned movement control codes. Group B had the smallest number of movement control codes. The two VR groups had the largest number of codes representing physical movement (turning in the chair with the VR headset), with Group D having the largest number of turns (130). Groups A and B both had much lower numbers of “rotating” view changes (29.25 and 24, respectively). The descriptives are available for each group in Table 6.

Table 5 - Average number of coded sections from each learner’s simulation use by group for two code types: Movement Controls and Planet & Information Details. Note: these are individual code instances, not the total amount of session time described by that code.

	A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Movement Control Code Sections (SD)	126 (8.63)	110 (39.21)	175 (65.54)	200 (20.68)
Planet & Information Details Code Sections (SD)	135 (30.44)	122 (28.23)	150 (47.03)	136 (25.16)

Table 6 - Average number of coded sections related to rotating the view (PC) or turning physically (VR) per group.

	A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Rotating Movement Code Sections (SD)	29.25 (12.4)	24 (20.46)	-----	-----
Turning Movement Code Sections (SD)	-----	-----	99 (37.96)	130 (12.97)

Chi-square tests were evaluated to examine differences between four overall categories between the four conditions: number of codes for planet and information details, movement controls, verbalizations, and overall rotating and turning (see Appendix K for all chi-square counts and details). There was a statistically significant difference between the number of codes, $X^2(9, N = 17) = 318.23$, $p < .001$. Standard residuals were evaluated to understand where the differences occurred. The two PC conditions (Groups A & B) had a higher than predicted number of codes for viewing planet and information details: Group A (z = 4.3), Group B (z = 5.03). In contrast, Group D had a much lower than expected number of visual viewing codes (z = -6.3). There were also differences between the number of verbalizations. Group A had a higher than expected number (z = 3.04), Group C had a much lower than expected number (z = -5.11) and Group D had a higher than expected number (z = 2.12). The VR groups had higher than expected residuals for the turning (which was an embodied, physical movement) compared to the PC Groups: Group C (z = 4.28) and Group D (z = 7.41); Group A (z = -7.83) and Group B (z = -7.64).

An additional exploratory chi-square analysis was completed for all four groups, looking for differences between codes: information panel open, viewing moons, viewing rings, and viewing the planet. A significant difference was found, $X^2(9, N = 17) = 233.66$, $p < .001$. A follow-up evaluation of the residuals identified where the differences lay. For both conditions in VR, there was a much less-than-expected number of information panel codes: Group C (z = -6.89) and Group D (z = -6.72). Group B also had less than expected (z = -1.99). Group D had the lowest amount of looking at moons (z = -2.98). The PC conditions had the lowest amount of zoomed viewing of the planets: Group A (z = -7.47) and Group B (z = -7.22). In contrast, Group D had a higher than expected amount of zoomed planet viewing (z = 2.54).

The types of verbalizations were evaluated between all four groups using a chi-square test and a significant difference was found, $X^2(9, N = 17) = 51.5$, $p < .001$. Standard residuals were tested to identify potential differences between groups. Those in Group A spent more time than expected verbalizing confirmatory facts (i.e., verifications for things they thought they knew) (z = 2.14). In the two VR groups, there were differences between the number of details shared, Group C had more than expected (*z* = 3.21) while Group D had less (*z* = -2.35). Finally, participants in Group D discussed more values (*z* = 3.77) and those in Group C discussed fewer (*z* = -2.17).

5.1.2 Emotional Engagement

PANAS is one standardized measure for positive and negative affect; positive affect includes alertness and enthusiasm and negative affect includes distress and other adverse moods. Higher positive affect may show more positive emotional engagement. There were no consistent differences between both conditions in the VR and PC groups (Table 4 B). Group D had the highest post-task positive affect score (43.5), but had a similarly low negative affect score compared to Group A.

Groups A and D had the lowest post-simulation use affect scores. Group D had the lowest variability in scores (1.3) and Group B had the highest (8.26). All groups had higher post-simulation use positive affect scores compared to negative. Differences between the post-use PANAS score for the PC and VR groups were evaluated using the Mann-Whitney U test, but none were found.

Table 4 B - PANAS post-use scores by group; higher scores represent high positive or negative affect.

		A) PC – no-audio	**B) PC – audio**	**C) VR – no-audio**	**D) VR – audio**
PANAS	**Positive (SD)**	31.8 (7.36)	30 (1.41)	28.4 (9.71)	43.5 (1.5)
	Negative (SD)	10.5 (0.5)	18.3 (8.26)	11.8 (2.23)	10.8 (1.3)

5.1.3 Cognitive Engagement

Four different measures were used to compare cognitive engagement during the learning activity: one standardized scale for evaluating engagement, accuracy of responses during the learning activity, and two measures for reflecting on cognitive engagement (post-activity solar system questionnaire and open-ended questions).

The SAQ includes six subscales: active engagement; superficial engagement; task mastery; ego/social orientation; work-avoidant orientation; affiliative goals (intrinsic motivation). Meece, Blumenfeld, & Hoyle (1988) identify active engagement and superficial engagement as the two components related directly to cognitive engagement. Active engagement includes strategies for self-monitoring of learning, while superficial engagement includes avoidant activity behaviors. Task mastery is an additional assessment of learner motivation, with a focus on interest in participating during the task and learning new content. These three subscales make up the cognitive engagement components of the SAQ (see Table 4 C for the overview). Scores were computed for all participants and averaged across groups. Each subscale has different maximum possible scores: active engagement is three; superficial engagement is three; and task mastery is four.

Example Superficial Engagement items included statements like “I skipped the hard parts” and “I guessed a lot so I could finish quickly.” Group A had the highest level of superficial engagement (2.4) while all other groups were 1.4 or less. Groups C and D had the lowest superficial engagement scores, at 1.24 and 1.2, respectively.

Example Task Mastery included statements like “I felt involved in my work” and “The work made me want to find out more about the topic.” Group D had the highest score for task mastery (3.91) compared to the other groups. Groups B – D had higher scores than Group A, with Group D over twice as high. Both VR conditions (Groups C & D) were higher than Group A (1.78). Group B had higher scores than Group A; Groups B and C had similar scores.

Table 4 C - SAQ subscales average scores by group; higher scores are better.

		A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
SAQ	Active Engagement (SD)	2.03 (0.52)	2.49 (0.14)	2.24 (0.45)	2.75 (0.2)
	Superficial Engagement (SD)	2.4 (0.51)	1.4 (0.47)	1.24 (0.08)	1.2 (0.35)
	Task Mastery (SD)	1.78 (0.32)	3.16 (0.64)	3.13 (0.64)	3.91 (0.10)

A Mann-Whitney U test was used to evaluate group differences between the VR and PC groups, and two significant differences were found. The PC group (Median = 1.9) had a significantly higher superficial engagement score than the VR group (Median = 1.2), $U = 16.0$, $p = .048$. The VR group (Median = 3.75) had a significantly higher task mastery score compared to the PC group (Median = 2.13), $U = 10.0$, $p = .012$.

Another measure of cognitive engagement could be how many questions participants answered accurately during the activity (Table 7), that is, were they able to interpret the questions, explore the simulation, and respond to them correctly, engaging with the material to accomplish a successful response? Learners in both the PC groups (A

& B) had a higher number of questions answered during the activity. Learners in Group C had the largest number of reported extra facts (i.e., details mentioned by the participant during the learning activity, which they observed while answering another question – this could include things like larger patterns in the planetary details).

Table 7 - Average number of during-activity questions answered and extra facts discussed by learner.

	A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Average Number of Questions Answered During Activity (SD)	7.75 (0.83)	9.25 (3.03)	7 (2.28)	4.75 1.16)
Average Number of Extra Facts During Activity (SD)	2 (0.71)	2 (1.41)	2.4 (1.02)	1.25 0.74)

When learners have changes in the number of correct responses (Table 8) and extra facts (Table 9) between the pre- and post-activity surveys, it may suggest higher cognitive engagement with the learning activity. In addition to the overall score changes, many often remembered supplemental details or knew further information than previously for portions of the more complex questions. All groups had increases in the number of correct responses from the pre-activity survey to the post-activity. Group D had the highest number of post-activity correct responses. Both Groups C and D had higher scores than the PC conditions. Group B had a higher score than Group A. There was no consistent increase or decrease in the average number of extra facts between the pre- and post-activity surveys. Groups A and C had an increase in reported extra facts, while Groups B and D had a decrease.

Table 8 - Average number of pre- and post-activity solar system survey correct responses reported by group; higher scores are better.

	A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Pre-Activity Correct (SD)	3.5 (2.29)	3.75 (2.39)	6.4 (3.88)	6.5 (1.34)
Post-Activity Correct (SD)	5.75 (2.86)	7.25 (4.21)	7.6 (2.58)	8.5 (2.86)

Table 9 - Average number of pre- and post-activity solar system survey extra facts reported by group; higher scores are better.

	A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Pre-Activity Extra (SD)	4.75 (1.64)	6 (3)	4.4 (1.2)	5.25 (1.16)
Post-Activity Extra Facts (SD)	5.75 (1.48)	4.25 (0.43)	5 (1.1)	4.5 (2.05)

The Mann-Whitney U test was used to evaluate differences between scores for the pre- and post-activity accuracy scores between the VR and PC groups. The VR group (Median = 7) had a significantly higher question accuracy for the *pre-activity* survey compared to the PC group (Median = 4), U = 15.5, p = .047; however, there were no differences between the *post-activity* scores or the overall score change for each group.

All participants also answered 6 open-ended questions at the end of their study session (these responses are all available in Appendix L). Participant responses gave insight to how they interacted with and interpreted information from the simulations. These questions asked about: what they remembered; what was hard to understand; what was easy to understand; what they liked; what they disliked; and what they wished they had more time with. Participant responses varied by describing things they <u>learned</u> (e.g., A201's "Venus is the hottest even though Mercury is the closest to the Sun"), things they did (e.g., B303's "I remember clicking on different planets and looking at and hearing different statistics for each"), or things they observed (e.g., C402's "Seeing all the planets and some of the weird orbits some of the moons had").

Twelve learners reported liking learning about and exploring planets they did not know much about (A202, A203, A204, B302, B304, C403, C404, C405, D501, D502, D503) or how real the experience felt (B303, C401, C404, D503). Seven learners (A: 1; B: 2; C: 2; D: 2) reported not disliking anything. One participant in Group A was sad they were not using VR for the activity, and one from Group B specifically said, "it seemed like it was over too quickly." Two participants in the PC conditions did not like User Interface/navigation settings (i.e., double-clicking to set the new focus); similarly, two in Group C disliked the movement (since it made it difficult to select a planet, "it was a little hard to click on the planets because the moons were usually in the way of it").

Every participant, except for A202 and B303, said they wished they had more time with the Universe Sandbox version. A202 thought, "No, it was enough time. It was all very straight to the point. It summarized the planet in a way that I could understand it." B303 reported, "I do not really wish for more time with Universe Sandbox because I feel that I have explored enough that I am satisfied." Seven participants, across all four conditions, wanted more time because they enjoyed the experience; in C403's words: "Yes, I had a really great time and I wanted to learn more about it. It was also very cool because you could really see the planets and it felt like you are really in space." Nine participants, again across all four conditions wanted more time; C402 explained, "Yes. I liked exploring and want to explore some more."

5.1.4 Motivation

Motivation is another important component which effects student engagement in learning activities. The MSLQ includes six subscales for measuring motivation: intrinsic and extrinsic goal orientation, task value, control beliefs about learning, self-efficacy for learning and performance, and anxiety (Pintrich et al., 1991, 2015; Pintrich & De Groot, 1990). Participants responded to these in the post-activity survey (Table 4 D). Small adaptations were made to the generalized scale (as the wording presented items in terms of "course materials" and "class activity") where necessary to reduce interpretation confusion for participants. Higher scores are better in all cases (maximum average score of 7), except for anxiety, where lower scores are better (minimum score of 1). Wording changes were discussed with pilot participants after they completed the study, to identify the incohesive item wording. Both the modified and unmodified questionnaires are in Appendix I.

Table 4 D - MSLQ subscales average scores by group; higher scores are better, except for anxiety where a lower score is better.

		A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
	Intrinsic Goal (SD)	3.88 (0.84)	5.31 (0.93)	4.55 (1.44)	5.63 (0.96)
	Extrinsic Goal (SD)	3.56 (1.72)	4.75 (0.90)	3.65 (1.64)	4.75 (0.88)
	Task Value (SD)	4.67 (1.26)	4.96 (0.82)	4.47 (1.4)	6.04 (0.59)
	Control Beliefs (SD)	4.69 (0.86)	4.94 (0.57)	4.25 (1.17)	5.63 (0.38)
	Self-Efficacy and Performance (SD)	4.97 (0.92)	4.59 (0.75)	4.55 (1.07)	5.09 (0.61)
MSLQ	**Test Anxiety (SD)**	3.3 (1.27)	5.05 (0.64)	3.04(1.21)	3.1 (0.83)

Pintrich, Smith, García, et al. state that high scores for Intrinsic Goal Orientation indicate higher student participation, or in other words, *student participation in the task is more than a means to an end* (2015). Example items included "In a learning activity like this, I prefer course material that really challenges me so I can learn new things," and "The most satisfying thing for me in this learning activity is trying to understand the content as thoroughly as possible." Group D had the highest overall intrinsic goal orientation score (5.63) compared to the other groups. Group C also had a higher score than Group A.

In comparison to the intrinsic motivation, extrinsic motivation (Extrinsic Goal Orientation) is a means to an end – *the learner had a higher motivation to finish the task, not to interact with the task but for other reasons* (Pintrich et al., 2015). Example items included "Getting a good score in this learning activity is the most satisfying things for me right now," and "I want to do well in this learning activity because it's important to show my ability." Levels of extrinsic motivation could be affected by additional factors like the learning tool used or the activity. The VR groups had higher extrinsic motivation scores than Group A. Groups B and D had equal extrinsic motivation scores, and Group B had a higher extrinsic motivation score compared to Group C.

Task value includes how interesting and useful the learners find the task, in relation to interest, importance, and utility (Pintrich et al., 2015). Example items included "It is important for me to learn the material in this learning activity" and "I like the subject matter of this learning activity." Group D had the highest average score for task value (6.042) compared to all other groups. Group C had lower scores for task value compared to the PC groups.

Control beliefs about learning relates to how learners perceive their effort as impactful to their learning experience outcomes (i.e., if they try hard, they might learn better). Example items included “It is my own fault if I don’t learn the material in this learning activity” and “If I don't understand the learning activity material, it is because I didn't try hard enough.” High or low control beliefs may affect the different ways in which learners interact with a learning tool. Group D again had the highest score for control beliefs (5.625) compared to the other groups, while Group C had the lowest overall score.

Test Anxiety is another component which had previously been shown to negatively relate to academic performance (Pintrich et al., 2015). Example items included “When I take a test, I think about how poorly I am doing compared to other students” and “I have an uneasy, upset feeling when I take an exam.” Test anxiety could affect how learners interact with new technology and influence their overall performance (with a decrement in performance if test anxiety is high), which is particularly relevant for activities such as a lab study, where they may believe they are being evaluated, even if the prompt is to “explore” or “discover.” Groups C and D had the two lowest Test Anxiety scores. Group B had the highest test anxiety score.

Each MSLQ subscale score was compared using the Mann-Whitney U test between the VR and PC groups; however, no differences were found between groups.

5.1.5 Long-term Engagement

Fourteen participants completed an optional follow-up phone call completed about one week after the study session (Group A: 3; Group B: 3; Group C: 5; Group D: 3). They all agreed that they would try a simulation similar to the Universe Sandbox again, and that something like this would be useful for school (Appendix M contains all responses from this semi-structured interview).

Across all of the groups, learners remembered details. Group D shared the most details about the planets in the open-ended questions (Table 10). A202 described a comparison between two planets, “I remember that, I believe it was Mercury and Mars had similar gravitational strengths.” C405 talked about the immersiveness, “It was like, kind of like being there, because I couldn’t see myself, but I could see all around, instead of just in one little spot. And like, the planets were moving and stuff.” For D502, the sounds for the rings were memorable: “Saturn’s rings, or any of the rings. Saturn’s stood out because it was both visual and sound.”

Table 10 - Average number facts remembered and reported by learners during the follow-up interviews when asked to describe the experience or a detail or two about the planets.

	A) PC – no-audio	**B) PC – audio**	**C) VR – no- audio**	**D) VR – audio**
Number of Facts Remembered (SD)	3.67 (0.47)	3.33 (1.25)	3.6 (1.63)	5 (2.16)

All participants reflected on the experience positively, and almost all of them thought it was very useful for understanding the distance between planets. B303 said, “It helped get it into perspective, but I kinda already knew.” Participants in the VR conditions had much stronger opinions: D503 said, “So, the simulation allowed, to show like, it made me able to see how these planets are actually a lot further or a lot closer together than I used to think they are. Like how the farther planets are a lot more spaced out.” C402 responded, “Yeah. I could see like, how big they were in relation to each other, or how *really* far apart they were;" and C404 agreed, "yeah, I think, having to teleport to each planet made me realize it was bigger, you couldn't just fly to them. Even the center planets weren't as close."

5.1.6 Qualitative Observational Results

Often, participants had qualitatively different experiences due to the assigned simulation condition. This section will focus on the individual experiences that are important considerations for any educational study, with anecdotes from the PC and the VR groups. In the PC group, some comments showed how immediately the learners grasped the idea of size and scale conveyed. For A204, in the first 20 seconds of simulation use, she said "Wow, you have to zoom out a lot to see, like, all of them." Afterward, she notes "It almost seems like the sun is like, kinda small, I mean, when you look at it as a whole thing." Others, like A201 observed interesting size comparisons partway through the activity, when he explored some of the moons: "Titan is tiny, but not as small as Mercury!"

A203 and A204 spent a great deal of time exploring the moons and observed the moon orbital patterns as they answered other questions during the activity. An exchange from A203 highlights how she completed this exploration:

> Experimenter: [Is there something going around Earth?]
>
> A203: "Moon?!"
>
> Experimenter: [So what do you think is going around Uranus?]
>
> A203: "Different types of like moons and stuff."
>
> Experimenter: [Are there a lot of planets that have moons?]
>
> A203: "Yeah."

> Experimenter: [Which ones?]
>
> A203: “The Earth, Uranus, I feel like that Mars had one, but I can’t check it” *then goes back to check it* “Yeah, Mars.”
>
> Experimenter: [Are there any ones besides Earth, Uranus, and Mars? Did Mercury or Venus?]
>
> A203: “No. No.” *explores* “Oh, Jupiter has some.”
>
> Experimenter: [Do any of the other planets have moons?]
>
> A203: *explores* “Saturn and this one, Neptune.”

After this exchange, at the end of the activity, this participant also spent the majority of the time exploring moons:

> Experimenter: [What are you checking out?]
>
> A203: “Moons. It’s fast, I can’t catch it.” *selects Io* “It’s around, like, Jupiter. The others, they look like blackish, they look similar, but this one, it’s colorful.” *moves to Ganymede* “This one is different too, it’s brown, but other moons don’t have craters, but the Earth’s moon does.”

A204 also carefully explored the orbital paths of multiple planets’ moons throughout the activity, both while answering questions about the planets generally (i.e., with the large info panel open) (Figure 8). She followed this type of search behavior consistently throughout the activity, even pausing to view the moons around Mars closely (Figure 9).

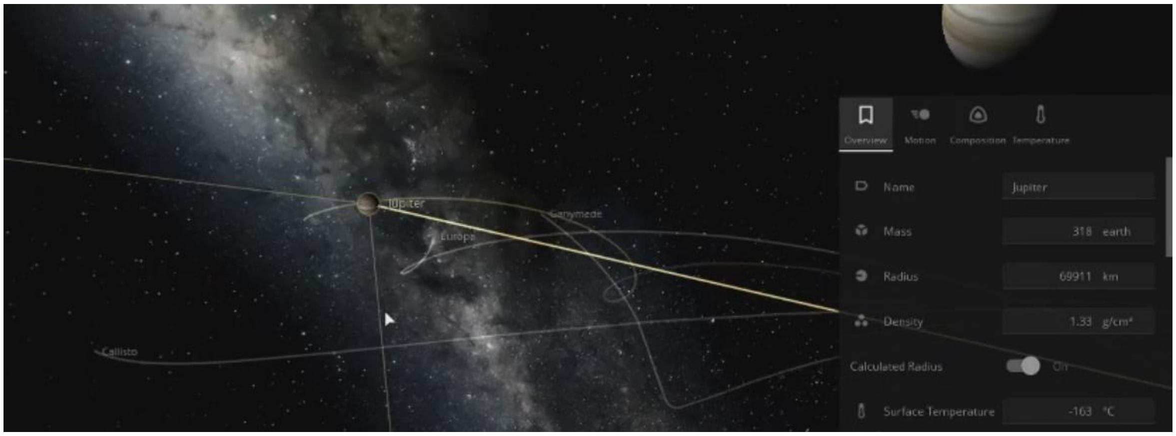

Figure 8. A204's zoomed-in view of Jupiter showing the main moons, while answering other questions about the planet using the large information panel.

Figure 9. A204 pauses the simulation to view the two moons orbiting Mars.

One common exploration pattern for PC included opening the large information panel and leaving it open as they completed multiple comparisons between planets. The length of time per participant varied, even within condition. For Group A, two participants left the large information panel open for more than half of the activity (A202: 10 minutes; A204 13.5 minutes); the other two participants barely used that larger information panel (A203 did not use it at all, and A202 opened it for only 14 seconds). The two participants who did not rely on the large panel instead left the small panel open and completed all comparisons that way. For Group B, different sets of exploration behaviors occurred. Both

B302 and B303 started exploring with the large panel open, during the middle of the activity relied on the small information panel, and then at the end returned to the larger panel. B301 open and closed the large information panel often throughout the activity, instead of leaving it open the entire time. B304 relied mostly on the small information panels for details.

Differences in movement control use also occurred within Group A. A201 utilized zooming twice as much as rotation for camera view changes, and followed that use pattern continuously through the activity. A202 used both rotation and zooming equally and consistently throughout the entire activity. A203 only began using rotation about seven minutes into the activity, and A204 used zooming constantly during simulation exploration, but only used rotation (about half as much) during the second half of the activity session. All four participants in Group B relied on zooming throughout the activity; however, B302 and B301 used zooming the most. B302 and B304 also relied heavily on rotation to change their view, compared to B301 and B303 who only sparingly used the rotation controls. B301 completed the most zoomed viewing of the planets for this group.

Even within the two VR groups, students explored the simulation differently. Some participants, like C404, paused the simulation movement at the beginning and did not restart it until much later. Since she started with it paused, she could easily teleport to a moon (like Oberon) and then quickly find the planet she was looking for (Uranus) afterward (Figure 10).

Figure 10. C404 finding Uranus after teleporting to Oberon, searching to her left, and then below her.

Other participants teleported to moons a lot (e.g., C401 and D501) but did not always pause the same amount, often times leading to them “chasing” a planet as the moon orbited. D503 had this exact experience (Figure 11): first, she teleported to Oberon, then she followed the orbit pathways from Uranus and its other moons, then as she chased down the planet and had trouble selecting it, was reminded about the pause button by the experimenter. Without this reminder from the earlier control introduction, she may have spent a large amount of time physically “chasing down” the planets when she teleported to a moon first. In future situations, she did use the pause button to reduce the amount of work it could have taken to otherwise travel to a planet. The total number of teleporting controls also diffed between VR participants, even though they all had the same orientation to the

VR controller. C402 completed the most teleporting movements during the activity (89), while four others (C401, C404, D502, and D503) had similarly high number of teleports (around 60), and four had much lower number of total teleports (C403, C405, D501, D504).

Figure 11. D503 chasing down Uranus after teleporting to Oberon; she spent over 30 seconds following the orbits before pausing the simulation.

An interesting case of mixed movement control use came from C403. He would point toward a planet, but if a moon label blocked the planet he wanted to teleport to, instead of teleporting to the moon and then to the planet, he would actually switch to the "Fly" mode and move closer to the planet to make it easier to teleport to it directly. D501 completed the most flying of any participant in the VR groups, and some (C401, C405, D503, and D504) did not use the fly movement control at all. Similarly, most participants did not use the grip button to pan the view, with the exception of C402 who used it quite often (over 35 times during the activity). Another behavior which varied between participants was resetting the simulation to the main view in order to re-orient: some barely used it, or did not use it at all (C402, C404, D502, D503); others used it some (C401; C405; D504); and a few used it quite often (C403; D501).

Learners in the VR condition relied heavily on the immersive, 3D environment to orient within the solar system. Instead of changing the view using controls like the grip button (which would have let them shift the view to a top down, or bottom up view of the solar system), learners would physically turn to search for planets, look for moons around a planet, or re-orient. In some cases, this helped learners find information they would have ignored otherwise; for instance, C401, when exploring which planets had moons said, “I don’t think Saturn has a moon,” then looked around some and corrected himself, “Wait a minute. I think Saturn might have a moon actually, Titan” (Figure 12).

Figure 12. C401 looking up to search for a moon around Saturn.

The previously described grip button view change did not occur often; however, one participant, C404, used the fly mode to achieve a similar view change which allowed her to look up at the orbital paths for each planet (Figure 13).

Figure 13. C404 looking up to view the solar system from below.

Although much of the differences for VR occurred between the types of movement controls and how learners explored the simulation, there were also similarities to the PC conditions. For instance, learners still focused on particular concepts or ideas and explored them thoroughly throughout the activity, even after they may have answered a question. One example of this occurred with a participant in Group D while he was exploring to find the hottest planet (Figure 14):

> D501: “Ok Venus right now is 475C, which is hotter than Mercury even though Mercury is closer to the Sun.”
>
> Experimenter: [Does that surprise you?]
>
> D501: “Yeah it does. I'm like Mercury is closer to the sun, it just feels weird to me. Maybe it's like, maybe the surface of Venus is different than Mercury. And that the surface is getting in a hotter way than Mercury."

Figure 14. D501 reflecting on the temperature differences between Mercury and Venus.

Often, learners would think that Mercury would be hottest because it was closest to the Sun, but if they checked other planets, too, would be surprised; D503 and D504 had similar experiences (D503 comments off-handedly about it, “Wow, this is much hotter”). In addition to these individual cases and examples, the Appendices contain individual responses to the open-ended questions after the activity (Appendix I), participant responses in the follow-up interview questions (Appendix M), and process code counts (Appendix K); all three of these provide more insight for the learner’s individual experiences with the simulations.

5.2 Research Question 2

RQ2. Does an audio-enhanced simulation or a visual-only simulation support higher levels of emotional, intellectual, and physical engagement?

Factors evaluated were cognitive engagement differences (post-activity scores, SAQ, follow-up interview responses), long-term engagement, usability scores (UMUX and

BUZZ, for the audio conditions), emotional engagement differences (post-activity PANAS), motivation (mATSI), and physical engagement (process coding for tool use and the Presence SUS). Similar to RQ1, the Mann-Whitney U test was used to evaluate differences between the audio and no-audio groups for each of the following factors: post-activity accuracy scores, UMUX scores, SAQ subscale scores, PANAS (positive and negative) scores, and SUS scores. Chi-square tests of independence were used to evaluate differences between the number of process codes assigned to each video to evaluate tool use and behavioral differences.

5.2.1 Cognitive Engagement

Group B had higher post-activity accuracy score on the solar system questionnaire than Group A (Table 8 A). Group D had a higher accuracy score than Group C, and the highest score overall. Pre-, post-, and overall change in accuracy scores (from pre to post) were compared to evaluate differences in content knowledge between groups.

The Mann-Whitney U test was used to evaluate differences between scores for the pre- and post-activity accuracy scores between the audio and no-audio groups. There were no group differences between the audio and no-audio conditions. No matter the condition, learners were equally successful and any prior differences were negated after simulation use.

Table 8 A - Average number of post-activity survey correct responses reported by group; higher scores are better.

	A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Post-Activity Correct (SD)	5.75 (2.86)	7.25 (4.21)	7.6 (2.58)	8.5 (2.86)

On the SAQ, the multimodal conditions had higher active engagement scores compared their partner no-audio conditions: Group B was 2.491 and Group A was 2.03; Group D was 2.75 and Group C was 2.24 (see Table 3C). Group B had a higher score than the Group C, perhaps an impact of the audio on increasing active engagement. The VR – audio condition (D) had the highest active engagement and lowest superficial engagement of all four simulation types. The audio conditions (B & D) also had lower superficial engagement scores than their no-audio counterparts (A & C).

There were also differences in the number of questions answered during the simulation-use activity: Group B learners had the highest number of questions answered during the activity (9.25), while Group D had the lowest number (4.75).

The Mann-Whitney U test was also used to evaluate group differences between the audio and no-audio groups. The audio group (Median = 2.56) had a significantly higher active engagement score than the no-audio group (Median = 2.0), $U = 15.0$, $p = .042$. The audio group (Median = 3.75) also had a significantly higher task mastery score than the no-audio group (Median = 2.38), $U = 14.5$, $p = .037$.

As mentioned previously, Group D participants reported the highest number of recalled facts, with some being from the sound layer in the follow-up interviews. There were no large differences between the other three conditions (Table 11).

Table 11 - Number of facts remembered and shared during open-ended question in the follow-up interview.

	A) PC - no-audio	**B) PC – audio**	**C) VR – no-audio**	**D) VR – audio**
Number of Facts Remembered (SD)	3.67 (0.47)	3.33 (1.25)	3.6 (1.63)	5 (2.16)

In the open-ended interview after simulation use, two participants in Group D disliked the sounds and wanted more time because, in D504's words, "I wanted to see if after a while I could understand the audio information." No participants in Group B reported negative feelings toward the sound layer.

5.2.2 Long-term Engagement

All six participants (from Groups B & D) who participated in the follow-up interviews mentioned the sounds when asked to describe the simulation and some details they remembered about it. B302, to answer the first question explained immediately that "There were different sounds that meant different things. The different things were like mass, or moons, rings, and other kinds of stuff for different planets about the solar system." This participant went on to describe what they remembered: "I thought it was pretty cool. Just like, going around space, hearing different sounds to know different things, and I liked not having to spend all of the time reading because I'm a slow reader. I remember being able to scroll out and see all of the planets, or scrolling in to focus on a single one. Like, when I was further out, I could hear all of the masses about the planets, but when I was in, I could hear just the one."

Those in the visual-only conditions were less likely to describe details unrelated to visual immersiveness. The eight participants from Groups A and C focused particularly on the movement of the moons and planets and did not recall as many other details: C405 explained, "It was like, kind of like being there, because I couldn't see myself, but I could see all around, instead of just in one little spot. And, like, there were planets moving and stuff." While these no-audio groups often recalled the movement, they recalled less specific details about the planets, with the exception of Group A participants, who reported more specific details: A202 recalls, "[I] remember that, I believe it was mercury and mars gravitational strengths were mostly the same."

5.2.3 Usability Scores

The Usability Metric for User Experience (UMUX) is a standardized scale for measuring usability (Finstad, 2010). Group B had a higher UMUX score than Group A. Group D had a higher UMUX score than Group C, and the highest score overall (21.75). Each score, and its analog percentage out of 100 is presented in Table 12. A Mann-Whitney U test was used to evaluate differences between the audio and no-audio groups. No significant differences in usability were found, even with the potential for complexity due to the additional layer of auditory information.

Table 12 - Average UMUX score per group. Higher scores are better, with a max score of 24.

	A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Average UMUX Scores (SD)	17.75 (1.09)	18 (3.94)	19.6 (2.94)	21.75 (1.09)
UMUX Percentage (out of 100)	73.96	75	81.67	90.63

BUZZ, the audio user experience scale (Tomlinson, Noah, et al., 2018), provided a measure of enjoyment and appeal as well as ease of use for the two multimodal conditions (Table 13). The original version of the scale has a high score of 77, across the 11 questions. Group B had a mean BUZZ score of 48 (SD = 9.427) and Group D had a mean BUZZ score of 56 (SD = 10.305). Group D had higher Aesthetics and Ease of Use Subscale scores than Group B.

Table 13 - Average BUZZ scores per group. Higher scores are better, with a max overall score of 77. The Enjoyment Subscale is out of 35, and the Ease of Use Subscale is out of 42.

	B) PC – audio	D) VR – audio
Average BUZZ Score (SD)	48 (9.25)	55.7 (10.31)
BUZZ Aesthetics Score (SD)	24.25 (4.32)	30.25 (3.27)
BUZZ Ease of Use Score (SD)	23.75 (5.40)	25.5 (7.92)

5.2.4 Emotional Engagement

Participants in Group A had the highest forecasted positive affect compared to those in Group B for the PC conditions. Group D had the highest positive affect for the two VR conditions. In the post-simulation use survey, participants in Group D still had the highest positive affect scores. Both Group A and Group B had an increase in positive affect. The two no-audio conditions (Groups A & C) had the highest variability in positive affect scores, and the audio conditions were lowest (around 1.5).

Group B learners had a higher forecasted negative affect score compared to Group A. Groups C and D had similar levels of forecasted negative affect. After using the simulations, all groups except for Group B experienced a decrease in negative affect. Group A had a lower negative affect than Group B, and Group D had a lower negative affect than Group C. All pre- and post-use PANAS scores are in Table 14.

The Mann-Whitney U test was used to evaluate differences between the pre- and post-activity positive and negative affect scores. No statistical differences were found between these groups.

Table 14 - Forecasted and post-simulation use PANAS scores.

		A) PC – no-audio	**B) PC – audio**	**C) VR – no-audio**	**D) VR – audio**
PANAS	**Positive – Pre (SD)**	31 (3.94)	28.75 (7.79)	33.4 (6.02)	41.75 (3.49)
	Positive – Post (SD)	31.8 (7.36)	30 (1.41)	28.4 (9.71)	43.5 (1.5)
	Negative – Pre (SD)	13.5 (2.29)	15.5 (2.87)	13 (2.19)	13.5 (2.96)
	Negative – Post (SD)	10.5 (0.5)	18.3 (8.26)	11.8 (2.23)	10.8 (1.3)

5.2.5 Motivation

Motivation between the visual and multimodal simulations was also compared using the MSLQ subscales (available in Table 4 D). Both audio conditions had higher intrinsic scores than their partner no-audio condition (D > C; B > A). Those in the multimodal conditions (Groups B & D) had higher extrinsic motivation compared to the no-audio ones (Groups A & C); they also had less variability in their scores. The multimodal conditions also had higher control belief scores compared to the non-audio groups.

Each MSLQ subscale score was compared using the Mann-Whitney U test between the audio and no-audio groups. The audio group (Median = 5.5) had a statistically higher intrinsic motivation than the no-audio group (Median = 3.75), $U = 15.5$, $p = .048$. No other differences were found between groups.

5.2.6 Physical Engagement

Between Groups A and B, B had the higher average SUS presence score (4.38). Within the VR conditions, Group D had the higher average SUS score (6.13). A Mann-Whitney U test was used to evaluate group differences between SUS scores for the audio and no-audio groups; however, no differences were found between them.

Process coding was also used to evaluate differences between simulation use and participant behavior during the study (Table 15). Group D had more codes than Group C and the most overall codes (427.25). Group A had more codes than Group B. Both VR conditions (Groups C & D) had more codes than the PC conditions (Groups A & B).

Group C had the most codes marking actions where learners viewed the visual/numerical representations of planetary data in the information panels. Groups C and D had the highest number of movement control usage while navigating in the simulation.

Group A had the highest number of non-task play actions. Group D had the highest number of verbalizations.

Table 15 - Average code counts for the four groups, representing the number of times that code was used to mark a section of the video. Note: these do not represent the total length of time the action/code marked quotations inside of each screen recordings.

	A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Total Code Count (SD)	301.25 (30.29)	265.5(66.09)	354.4 (111.19)	427.25 (81.8)
Planet & Information Details Count (SD)	135.25 (30.44)	122 (28.23)	149.8 (47.03)	136.25 (25.16)
Movement Controls Count (SD)	126 (8.63)	109.5 (39.21)	175.4 (65.54)	200 (20.68)
Non-task Play Count (SD)	8 (3.03)	0.25 (0.43)	1 (0)	1 (0)
Verbalizations Count (SD)	31 (37.25)	27.75 (25.78)	25.4 (16.65)	53 (34.78)

The two audio conditions (Groups B & D) had lower interaction numbers with viewing the planetary information detail panels, which was likely due to the additional auditory representations (see Table 5).

Chi-square tests of independence were used to evaluate the observed number of process codes per condition compared to the expected, to test differences between conditions. Differences between Groups A and B were compared for information panel tab use, including use of individual tabs (e.g., Overview Tab, Climate Tab) and also for longer periods of use for the small and large information panels being open across multiple planet comparisons. No significant differences were found between the audio and no-audio PC conditions.

A chi-square test was completed to compare between the two VR conditions evaluating usage differences for the movement controls (e.g., teleporting, flying) and orientation information (e.g., showing planet labels). Significant differences were found between the VR groups, $X^2(8, N = 9) = 70.7$, $p < .001$. Residuals were evaluated to determine where the between group differences occurred. Participants were more likely to orient in the no-audio condition (Group C) ($z = 2.54$) and less likely to orient using visual labels in the audio ($z = -2.69$). The no-audio group was also more likely to explore using the grip button to change the view ($z = 2.07$) and the audio group was less likely to use the grip button ($z = -2.26$). Group D also teleported to the sun less than expected ($z = -2.08$). Physical or embodied movement also occurred less in the VR – no-audio group ($z = -2.52$) and more in the VR audio group ($z = 2.76$). Finally, participants in the no-audio group paused the time less ($z = -3.01$) and audio group paused time more ($z = 3.3$).

5.2.7 Qualitative Observational Results

Participant experiences qualitatively differed between the audio and no-audio groups as much as they differed between the PC and VR groups. Learners relied on the sound and talked about the simulation data with respect to the sound as they explored and answered questions. For instance, of the participants in Group B, B302 discussed the audio representations the most, and he used previously-known audio vocabulary (e.g., pitch) when discussing how he was interpreting the audio. One example comes from his exploration of temperature (Figure 15):

B302: “Explore and discover what the two coldest planets are. I’d think they’d be the farthest out ones. So… it’s like based on how high and low their waves are, right?”

Experimenter: [On the climate panel]

B302: “Right.” *opens the climate tab* “Hmm…”

Experimenter: [What do you notice about that one?]

B302: “It’s pretty quiet, so that’s pretty cold. Neptune” *selects Neptune* “I also don’t really hear anything. So that’s also cold. Next one would be Saturn” *selects Saturn* “I hear something from there so it’s warmer. Jupiter” *selects Jupiter* “Hear something.” *selects Mars* *selects Earth* “Venus” *Clicks Venus* “And Mercury” *Clicks Mercury* “So, Uranus and Neptune.”

B302: “Explore and discover what the one hottest planet is. So, I can automatically rule out Neptune and Uranus cause I know they’re the coldest. I believe I heard that Venus was higher pitched than Saturn, but checking… It’s pretty quiet.”

Experimenter: *After some time of the student exploring* [So what are you thinking about?]

B302: “Just really trying to compare if Venus or Mercury is higher pitched.” *compares the numerical values* Definitely Venus.”

Of the PC – audio group, B302 spent the most time purposefully exploring with the Climate and Overview Tabs open. Certain participants seemed to focus on particular types of sonification while they explored. B301 mostly relied on the Overview Tab and barely explored the others during the activity. B303 relied most on the Overview and Motion Tabs, while B304 spent the most time exploring the sounds related to rings.

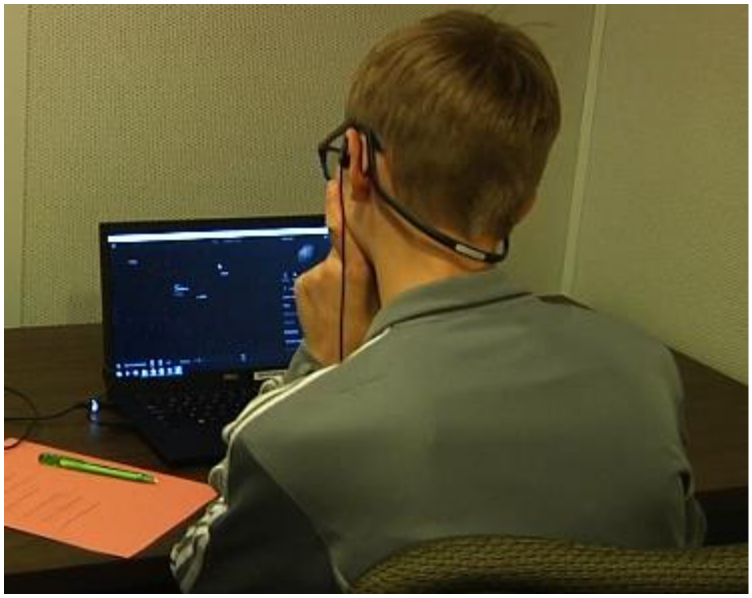

Figure 15. B302 exploring to find the coldest planets; the climate tab is open to play the temperature sonification.

Learners in Group D also relied heavily on the sounds to answer many of the questions. D502, while looking for planets with shorter and longer lengths of day than Earth, discussed the sounds and moved his body in time with them (Figure 16):

> Experimenter: [Explore and discover a planet with a day longer than Earth's.]
>
> D502: "A day longer than Earth's? Ok." *first goes to Uranus, then goes to Earth* "Start at Earth. Maybe I should start with Earth, just for comparison." *listens to the sound while moving his hand up and down to the sound a few times, then shakes his head up and down in the same pattern*
>
> D502: *teleports to Mercury* "So compared to Earth's beat this one is more of a drag. So.."
>
> Experimenter: [So, what do you think that means?]
>
> D502: "A drag is probably a longer day." *teleports to Venus* "We'll go to Venus. Also a drag, so probably a longer day on Venus as well." *teleports to Mars* "Mars is almost the same as Earth's, I'd say." *teleports to Saturn* "Saturn seems to be faster than Earth." *teleports to Jupiter* "Jupiter also seems to be faster, so probably that for Uranus and Neptune as well." *goes to Neptune* "Neptune actually seems slower, but still around the same speed as Earth. And finally, Uranus" *teleports to Uranus* "A little slower."

Throughout the exploration, D502 consistently moved his hand and head in time with the sounds to help make comparisons between different values for each planet.

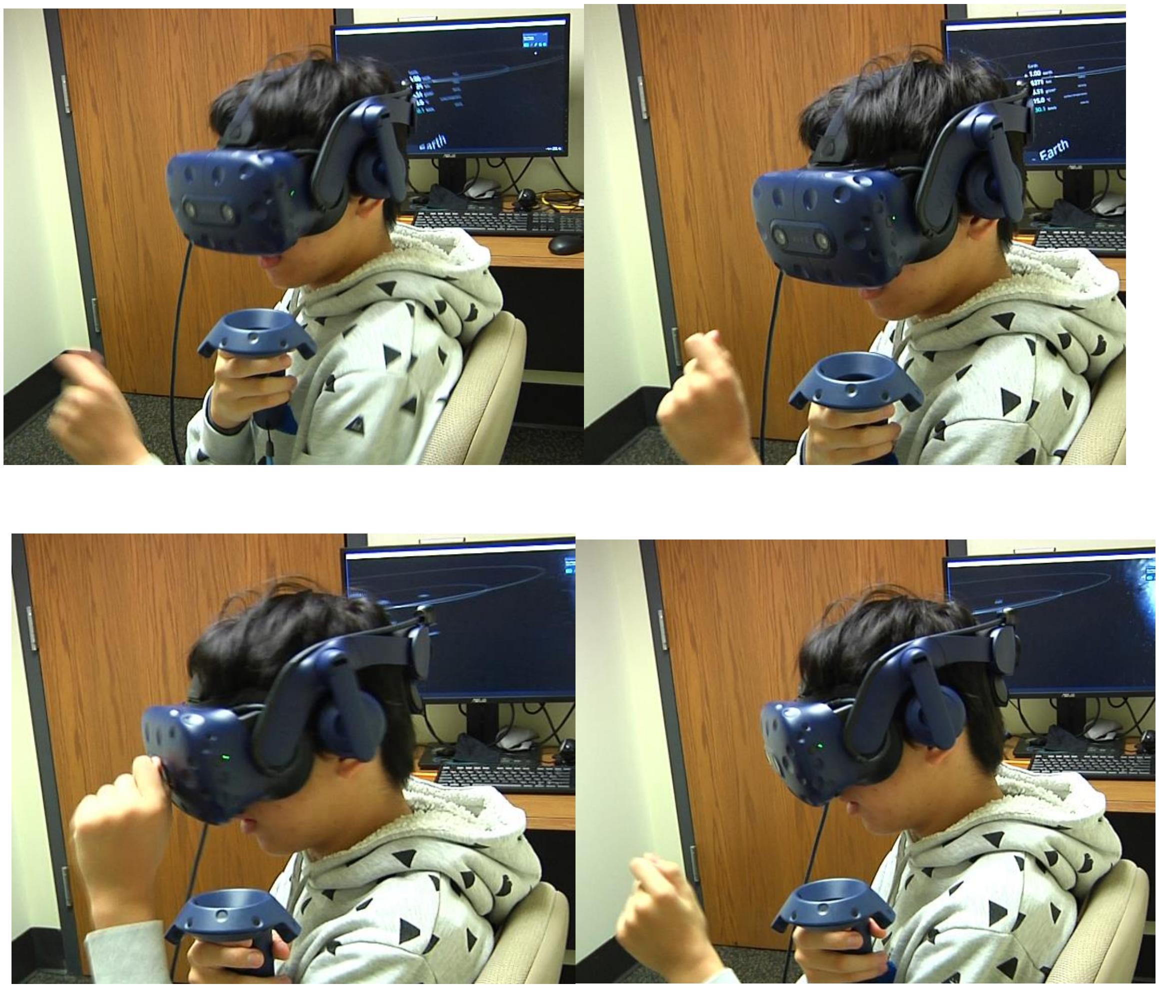

Figure 16. D502 moving his hand up and down while listening to the mapping for length of day on Earth.

Not all participants in Group D discussed the sounds as much while they were exploring. Three of participants (D501, D503, and D504) talked about the sound mappings the most while searching for planets that had rings, and would re-teleport to a planet to verify whether or not that sound occurred. This is just a subset of experiences from the learners and the Appendices have more details which sounds the participants recalled in the open-ended questions and follow-up interviews.

5.3 Research Question 3

RQ3. What factors, such as technology experience, math and science anxiety, self-efficacy, and affect, influence a student's ability to interact comfortably with multimodal science tools?

Factors evaluated to understand their influence on new technologically supported learning experiences included technology use, forecasted affect (PANAS), additional science activities, and learner perception towards science/anxiety/self-efficacy (mATSI). The Mann-Whitney U test was used to evaluate differences between the VR and PC groups and the audio and no-audio groups for each of the following factors: forecasted PANAS (both positive and negative), technology use scores, and the mATSI subscale scores.

5.3.1 Technology Use

The Adapted Technology Experience Profile (Liles, 2017) was used to measure learner experiences with the 20 listed technologies. This questionnaire was updated for students (Liles, 2018), since the original profile was created for adults (Gonzalez, E. T., Mitzner, T. L., Sanford, J. A., & Rogers, 2016). Participant's frequency profile score and general breadth profile scores were calculated following Barg-Walkow, Mitzner, & Rogers (2014).

Across all participants (Table 3 A), the frequency profile score (or how often they used the entire set of 20 technologies in general) was 1.26 (SD = 0.36), out of the maximum score of 3. For the general breadth of technology score, the average was 10.47 (SD = 2.62) out of a maximum score of 36.

Five categories related specially to categories which may affect simulation use (experience with PCs, video games, PC games, Augmented Reality headsets, and VR headsets). The average "game use" frequency profile score was 1.34 (SD = 0.52) out of 3, while the average breadth profile score was 2.84 (SD = 0.98) out of 5. Group B had higher general and more frequent technology and game usage compared to Group A. Group D had more frequent technology usage than Group C, although the general technology usage was the opposite.

Table 3 A - Technology use Frequency Profile and Breadth scores by group; higher scores represent more use.

		A) PC – no-audio	**B) PC – audio**	**C) VR – no- audio**	**D) VR – audio**
Frequency Profile	**General** (Out of 3)	1.01 (0.28)	1.07 (0.18)	1.44 (0.28)	1.47 (0.22)
	Game Use (Out of 3)	0.95 (0.17)	1.3 (0.36)	1.44 (0.41)	1.55 (0.26)
Breadth	**General** (Out of 36)	8.25 (2.49)	9 (2.55)	12.2 (1.47)	12 (0.71)
	Game Use (Out of 5)	2 (1.41)	2.5 (0.03)	3.4 (0.49)	3.25 (0.43)

Between group differences on technology use were compared for the VR and PC conditions and audio and no-audio conditions using the Mann-Whitney U test. The VR group (Median = 1.53) had a higher general frequency profile score than the PC group (Median = 1.08), $U = 10.5$, $p = .014$. The VR group (Median = 12) also had a higher general breadth of technology use score compared to the PC group (Median = 8.5), $U =$

10.0, $p = .010$. That same group (Median = 3) also had a higher breadth of relevant game use compared to the PC group (Median = 2), $U = 13.5$, $p = .020$.

5.3.2 *Forecasted Affect*

Although PANAS was created for reporting past (or current) affective states (Watson et al., 1988), researchers have more recently used it, with descriptive passages, to help *predict* how users feel about certain situations before other activities take place (Noah et al., 2016). Calderwood, Green, Joy-Gaba, & Moloney (2016) have also used it to predict leaner affect and number of errors using a variety of multimedia educational tools; they found students accurately forecasted the negative affect effects from multimedia intervention.

Each group answered the PANAS to forecast affect before simulation use (Table 3 B); there was a specific prompt for each simulation version. For example, learners in Group D had this prompt:

> This scale consists of a number of words that describe different feelings and emotions.
>
> Imagine you are going to use an interactive simulation with sounds on a virtual reality headset. This simulation will be used to display information about our solar system.
>
> Thinking about what this experience may be like, to what degree do you think you would experience each of the following feelings or emotions while using the interactive virtual reality simulation with sounds.

Group D had the highest positive predicted affect, while Group B had the lowest. Groups A, C, and D had similar predicated negative affect (around 13), while Group B had the highest predicated negative affect. Group differences in forecasted affect were tested between the VR and PC and audio and no-audio groups, although no statistical differences were found.

Table 3 B - PANAS pre-simulation use scores by group; higher scores represent higher positive or negative affect.

		A) PC – no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
PANAS	Positive (SD)	31 (3.94)	28.75 (7.79)	33.4 (6.02)	41.75 (3.49)
	Negative (SD)	13.5 (2.29)	15.5 (2.87)	13 (2.19)	13.5 (2.96)

5.3.3 Additional Science Activities

Other factors which may influence a student’s learning experience could be predicted from their prior experiences in other types of interactive learning environments. These may include informal activities such as museums, zoos, aquariums, planetariums, and science centers. The frequency with which they attended these may predict how positive they might feel about the exploration activity beforehand. There were some differences between the number of informal science activities, with the VR group participating in a few more than the PC group (Table 16).

Table 16 - Reported attendance of types of Informal Science Activities by group.

	A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
Average Number of Informal Science Activities (SD)	1.75 (0.43)	2.5 (2.06)	3.2 (1.33)	3.75 (1.64)

Additional activities which may predict higher forecasted affect scores: following any science/news media; attending STEM summer camps; or previous class projects on the topic of space. More participants in Groups B and D followed science news or other media outlets regularly, with Group A having the fewest number (Table 17). Eleven of the seventeen participants had attended a STEM summer camp, and thirteen of them had completed some project about space for school. These participants were distributed evenly through the groups.

Table 17- Reported Science/News Media (including TV shows like NASA, Discovery, or online news from Twitter, Facebook, etc.) following by group.

	A) PC - no-audio	**B) PC – audio**	**C) VR – no-audio**	**D) VR – audio**
Average Number of Science/News Followed (SD)	1.25 (1.3)	2.75 (0.83)	2.25 (1.41)	3.75 (1.3)

5.3.4 Learner Perception Towards Science

The mATSI is a questionnaire developed to measure factors which influence student learning, including perception of the teacher, anxiety toward science, values of science to society, self-confidence in science, and desire to do science (Weinburgh & Steele, 2000). The four relevant subscales were evaluated (Table 3 C). Group differences were evaluated for each mATSI subscale between the VR and PC and audio and no-audio groups; however, no statistical differences were found.

Table 3 C - mATSI Subscale scores. Higher scores are better, except for anxiety subscale (lower scores are better). Total scores: Value of Science to Society = 25; Anxiety = {min = 5; max 25}; Self-Confidence in Science = 25; and Desire to Do Science = 35.

		A) PC - no-audio	B) PC – audio	C) VR – no-audio	D) VR – audio
mATSI	**Value of Science to Society (SD)**	18.75 (4.32)	20.5 (3.28)	21.2 (1.47)	22 (0.707)
	Anxiety Towards Science (SD)	9 (2.74)	8.75 (2.49)	9.8 (3.71)	7.5 (1.80)
	Self-Confidence in Science (SD)	19.75 (1.92)	21 (2.12)	21 (2.61)	20.75 (2.59)
	Desire to Do Science (SD)	24.25 (3.56)	26 (4.24)	24.2 (6.43)	29.75 (3.45)

CHAPTER 6. STUDY 3 DISCUSSION

Study 3 explored specific differences between four simulation versions, and evaluated their ability to support learning opportunities. It tested different factors which may affect positive or negative learning experiences, including engagement, motivation, usability, and ease of use; it also considered the influence of prior science and technology experiences.

It was predicted that the VR conditions would have higher levels of engagement compared to the PC conditions, particularly in the physical and cognitive engagement scores (evaluated for RQ1). It was also predicted that the multimodal (or "audio") conditions would have higher levels of engagement than the visual only ("no-audio") conditions, based on the Multimedia Principle (evaluated for RQ2). Factors like prior experience with technology and engagement with science through a variety of aspects were evaluated to understand any potential impacts on technology use (evaluated for RQ3).

6.1 VR and PC Differences

Simulations are used in learning to support student exploration of models, systems, and content which may be difficult to interact with or observe otherwise due to size, scale, and complexity. To evaluate the ability for the simulation to support a student's opportunity for learning, a pre- and post-activity solar system questionnaire was given to participants. Despite initial significant differences between participate knowledge on the pre-activity questionnaire – the VR group had a higher score than the PC group – after the simulation activity there were no statistical differences. Both the VR and PC conditions increased student content knowledge, and the VR conditions increased more even though participants completed less questions during the exploration activity. The VR simulation provided an immersive environment where the learners could build contextual and periphery knowledge as they explored, leading to higher overall accuracy scores on the post-activity questionnaire (aligning with previous results from Slater & Sanchez-Vives (2016) about exploration outside of real-life constraints). Previous research has also supported the ability for VR environments to be more inclusive, extensive, surrounding, and vivid (Slater & Wilbur, 1997); VR participants reported these kinds of details in their follow-up interviews.

Many other validated scales and measures were used to compare different types of engagement (cognitive, physical, emotional) and motivation. The VR groups had higher motivation for participation in the task (task mastery) compared to the PC group, showing differences in the potential for VR to enhance the learning experience past what is typically available from a computer simulation. These differences were also reflected in the higher MSLQ motivational scores measured after the simulation use.

In addition to the higher motivation, the VR groups had a significantly lower superficial engagement score than the PC groups. This was also reflected in the higher active engagement scores. Taking these two engagement scores together suggests that the VR simulation supported better cognitive engagement than the PC, perhaps due to the immersive environment which surrounded the learner and provided contextually situated information. One example of this was how far away planets actually were when they tried to fly between them, even the ones that are closer to the sun. Many learners in the VR commented on this: C402 explained, “Yeah. I could see like, how big they were in relation to each other, or how *really* far apart they were;” and C403 agreed, “Yeah, I think, having to teleport to each planet made me realize it was bigger, you couldn't just fly to them. Even the center planets weren't as close.”

Comparing behavioral differences through the process coded video recordings gave additional insight toward participant experiences. For example, when the PC group verbalized (and compared) details they were confirming prior knowledge more than any other group. While this is a positive behavior in general, the VR groups still discussed details and values for the planets more than either of the PC conditions, even though the PC groups had more details available through the large tab panels.

During the study, participants in the PC conditions were able to modify the default simulation settings (e.g., changing the sun’s mass), and did so, particularly in the free play period at the end of the activity. This may have contributed to the overall higher scores with respect to their empowerment and self-efficacy for this particular learning activity, compared to the no-audio VR group. With that logic, the expectation would be that the VR – audio learners would also feel less empowered about their learning. However, this group felt more empowered about their learning experience than any other group, possibly due to their ability to explore planetary details in multiple ways through multiple representations (visual and audio). If students feel that they can directly influence their learning outcomes or academic performance, they are more likely to exert themselves actively in the activity (paralleling Ryan and Deci’s notion that supporting competence and autonomy lead to better motivational maintenance (2000)). As these results were mixed, future work should explore the influence of technology type on student motivational factors.

There were no differences in emotional engagement after simulation use, so even if novelty of technology could influence a learner's initial experience, it may not be significant enough to drive continued use or engagement. Additionally, it is possible that differences in emotional engagement may have occurred during simulation use but were reported differently retrospectively (a pattern which has some precedence in previous research (Matsumoto & Sanders, 1988)). Including small queries to explore participant's affect *during* simulation use may provide additional insight into differences in emotional engagement.

Finally, statistical differences were observed for physical engagement through both presence scores (SUS) and process coding for the VR groups. Higher presence scores suggest a more immersive learning environment; this immersion may lead to longer-term positive learning opportunities, particularly for visual representations such as size or distance. Through the coding, the VR group had a much larger amount of turning to orient, observe, or explore the simulation compared to the PC group. The VR – audio group had the most zoomed observation of planets compared to both of the PC conditions. That group discussed the most values even though they looked at them less frequently, and did not focus as much on confirmatory exploration the way PC – no-audio group did. VR supported different exploration behaviors (e.g., less focus on numerical visuals presented through the information panels), which lead to higher overall post scores, physical engagement, and high scores for positive affect, motivation, and cognitive engagement.

6.2 Audio and No-audio Differences

The Multimedia Principle has historically shown that students have better learning outcomes when using a combination of animations and narration (or also simply, pictures and text descriptions); however, there is a dearth of research comparing the effect of animations, non-speech audio, and text (and even just visuals and non-speech audio). Sonification research has evaluated the use of non-speech sounds for supporting learning opportunities for students with impairment, but similar systems are not widely studied in the context of education literature. Study 3 compared participant experiences in these audio-enhanced learning environments to begin unpacking non-speech audio's potential for supporting the Multimedia Principle.

Learners in the multimodal simulation conditions had higher post-activity questionnaire scores, though they were not statistically different. The VR – audio condition had the highest overall score and the PC – audio condition was higher than the PC – no-audio. Participants in the PC – audio condition had the largest increase in post-activity scores, and also the largest number of answered questions during the activity. The detailed descriptions available in the PC condition, as well as the additional information available through the audio content, may have impacted these learners the most.

The audio conditions had significantly higher intrinsic motivation (MSLQ), active engagement (SAQ), and task mastery (SAQ) than the non-audio groups. The PC – audio group also had a higher intrinsic motivation than the VR – no-audio group, providing some evidence for the multimodal component resulting in higher motivation scores overall, compared to just the introduction of a new technology. Adding multimodal or audio components could be particularly useful for learners who have lower levels of motivation, and would not otherwise engage with the materials as deeply as learners with a high intrinsic motivation.

Active engagement is necessary for learning activities to be successful, and higher active engagement ties into higher cognitive engagement for learning. The audio conditions had statistically higher scores for active engagement compared to the no-audio conditions, possibly due to the representation providing an additional layer of information for learners to engage with during the activity. The two audio groups also had lower superficial engagement compared to their no-audio counterparts, suggesting additional positive benefits from the embedded auditory displays.

Task mastery involves interest in participating during the task and learning new content, a behavior which should be high for a learning activity to be effective. The audio layer presented additional information for the learners to explore, and gave them more opportunity to be motivated to interpret and understand the represented simulation content.

Another important component for engagement is emotional engagement. No significant differences were found between the PANAS scores for the audio and no-audio groups; however, the descriptive differences were large and of interest here. The VR – audio group had the highest positive affect after simulation use. Learners may have really enjoyed the experience and using a multimodal VR simulation turned out better than they thought. This group had lower negative affect scores compared to both its no-audio counterpart (C) and the multimodal PC condition (B).

Group differences for affect scores occurred between the PC conditions, too. The multimodal PC condition had a lower positive affect and a higher negative affect compared to the visual-only condition. Two different participants in the PC – audio group may have contributed to this difference. The first had the highest negative affect score (30) – which inflated the overall group average. This participant (B301) had lower score on the pre- and post-activity solar system questionnaire; they also had lower scores for accuracy in responses during simulation use: two correct out of 11 answered questions. In the open-ended questions, B301 reported being unsatisfied and wished they had more time "because it was fun." If B301 had more time with the simulation, they may have reported different emotional engagement scores.

Participant B303 also had a higher negative affect score (22) compared to the typical range (10-15) of other participants. This learner had a different experience with the Universe Sandbox activity: they only answered 4 questions using the simulation, but still had an increase in accuracy score from the pre- to post-test. When asked, B303 felt they had enough time with the Sandbox "I feel that I have explored enough that I am satisfied." The total impact of audio on emotional engagement is unclear, and should be studied further in future research.

Learners in the VR – no-audio group also had a lower positive affect; in the post-activity open-ended questions, three Group C participants (C402, C403, C405) all reported that it was a slightly difficult to contextualize where the planets were when not in the top-down view, and that the moons moved really quickly. Both of these settings could be easily addressed through more familiarity with the movement controls, and seemed to be unrelated to the presence (or absence) of audio. Comparing the code counts, VR – audio participants completed more movements (including resetting the view) – all things which would have impacted their ability to remember the context and layout of the planets as they moved. The VR – audio group also had more pause/time scale changes than its no-audio counterpart, maybe due to stopping at planets to listen to the auditory displays. This may account for their more positive experience and reduced frustration compared to the non-audio group.

Audio user experience (BUZZ) was another factor compared between the two audio groups. Both groups had the same auditory representations embedded within the simulation, although they were accessed differently depending on the UI controls. The PC – audio group had more visual numerical representations for details when the large panel was expanded. Potentially, the simplified visual representations available in the VR – audio condition was easier for learners to integrate into their exploration experience, and that was reflected through higher ease of use scores. Higher aesthetics scores could be due to the spatial nature of the planet's mass sounds coinciding with the visually immersive 3D environment. Although the sounds were spatialized in the PC – audio condition it may not have had the same immersive effect for those learners.

The audio conditions (Groups B & D) had higher presence scores than their non-audio (Groups A & C) counterparts, although these were not statistically different. Additional differences between the VR conditions were revealed through the process coding, and Group D had more physical or embodied exploration than the non-audio group, showing the potential for audio or multimodal simulations to support more physical engagement than visual-only.

The non-speech audio did lead to differences in learner experiences, including significant impacts on cognitive engagement and motivation. Participants in the audio conditions explored differently, had more physical engagement, and spent less time orienting, particularly in the VR audio condition. The PC audio condition participants also spent less time viewing the numerical values compare to the regular PC condition. Audio seems to provide additional means for engagement, and alternative information to focus on during the activity, instead of learners completing non-learning-activity related exploration (e.g., exploring the Sun or verifying content they already knew).

6.3 Practical Implications for Educational Simulations

Teachers, administrators, and researchers might have concerns about using novel (due to either technological or modality differences) simulations for student learning experiences. One potential impacting factor would be usability: does introducing a new modality type or technology into the equation lead to usability challenges? From the Study 3 results, there were no significant differences in usability for either the VR conditions or the multimodal conditions; neither resulted in complexity which interfered with participant learning opportunities compared to a typical simulation experience.

Others may be concerned over introducing non-speech audio as a secondary information layer: could this distract learners from the visual information? These data show that they did not distract them. In fact, the VR – audio group paused more, relied less on teleporting to the sun and using visual labels to orient, and they moved more than the VR – no-audio group. They even had more exploration of values than the VR – no-audio group, even though they viewed the info panel less. The PC – audio group also looked at the information details less than the PC – no-audio group.

As mentioned previously, in the follow-up interviews, Group D participants reported the highest number of recalled facts, with some details coming directly from the sound layer. There were no large differences between the other three conditions, so the combination of the VR and the sound layer may have contributed to Group D's higher number recalled facts.

The audio led to significantly higher cognitive engagement than the no-audio conditions, and again, the VR – audio condition had the highest level of active engagement. Interestingly, the PC – audio version (B) had a higher engagement level than the VR – no-audio (C), suggesting that focusing on visual-only VR may not be enough to support the best learning opportunities.

Others have brought up concerns that additional factors like technology experience, anxiety and self-efficacy may affect learners comfort engaging with newer technologies for learning. Although both all groups were randomly assigned, participants in the two VR conditions had statistically significantly more frequent experience with the 20 technologies listed, more frequent experience with VR or game technologies, and a slightly larger breadth of exposure to technologies, though all groups had similar experiences with trying game-related technologies. Although these differences existed beforehand, they did not seem to lead to any statistically different levels for things like post anxiety scores, emotional engagement, or overall differences in facts remembered or accuracy scores. This suggests that even with little prior experience, students can adjust and be comfortable using new technology for exploratory interventions. It could also mean that the novelty effect, which usually has little long-term motivational influence on students, may not have influenced the participants' experiences (for more information on this novelty motivation, see Liu, Toprac, & Yuen (2008)).

Other factors may predict differences in overall pre-activity accuracy scores or post-activity reported experiences. Comparing the group averages for the types of informal learning activities they attended, participants in the VR conditions also attended slightly more informal learning activities. This does coincide with their higher forecasted affect for those two conditions before using the simulation (though it was not significantly different). Many of the learners did follow other forms of science news, participate in STEM summer camps, or study space for school projects. Each of these three factors may have influenced learners, perhaps with the overall high participation, high interest, and high technology experience adding to the VR group's higher positive affect scores both before and after the experience. The influence of these factors should be explored more with a larger sample, to determine whether they have an affect or not; this study suggests there may not be a large effect.

Interestingly, the VR – no-audio condition participants had the highest reported mean score for Anxiety Toward Science (9.8); it remains unclear what might account for this forecasted affect difference. Learners in this group had the highest participation in additional activities like summer camps. They also had higher participation in other news/media about science when compared to those in the PC – no-audio condition, even though they actually had much higher anxiety scores towards science. There were no significant differences between the number of extra informal science activities and the number of external science news followed on social media and news outlets. Targeted informal education interventions may lead to differences (Bonney et al., 2009); however, it seems as if there was no difference on this particular simulation experience when considering these factors.

These higher anxiety scores from Group B may also account for their lower overall post-simulation use PANAS positive scale scores. If these learners were more anxious about science in general, the positive influence of the audio may have not had a higher positive affect, compared to the VR – audio experience. Although, given this group's lower post negative affect score, the experience may not have been as bad as they predicted it would be. This should be studied further in the future to better understand the impact of long-term use of technologies like simulations, and whether or not they can help change learner perceptions about science anxiety.

Self-efficacy is another factor which typically predicts outcomes of educational experiences (Pintrich & De Groot, 1990). All learners had similar confidence about their self-efficacy in science before completing the activity (mATSI – Self Confidence in Science subscale ranged from 19.75 to 21 for the four groups, with no statistical differences). In the post-simulation use surveys, all participants again had similar self-efficacy and performance scores on the MSLQ's measure, with no significant differences. When reflecting on their experience after using the simulation, many of the learners thought back on their experience positively; for example, A203 said, "I liked that I got to explore the planets that I didn't know." Each participant could explain factors they liked about the system, including the ability to view planetary information. Hopefully, over time, simulations and other approachable science tools may help increase self-efficacy for science more generally.

Did this translate to positive experiences after the activity? All participants answered a series of 6 open-ended questions. One question asked “Did you wish you had more time with Universe Sandbox? Why or Why Not?” This was used to evaluate whether or not learners enjoyed the experience, and if they found it to be worthwhile. All participants, except two, said they wished they had more time. The two who said “no” shared similar opinions (stated here by B): “I do not really wish for more time with Universe Sandbox because I feel that I have explored enough that I am satisfied.”

All groups had a positive experience upon reflection on their simulation use. Many participants wanted more time using the simulation to explore and learn more; even learners who had trouble with interpreting some of the audio representations wanted more time to explore them. Slightly more participants in the VR conditions (Groups C & D) reported how they enjoyed learning about the planets and how real the experience felt. Equal numbers of participants in the sound conditions (Groups B & D) and the VR conditions (Groups C & D) reported they did not dislike anything. Flexibility of exploration, presented details, and the additional auditory information may have contributed to this shared opinion.

All groups had a positive response when reflecting on the learning experience during the follow-up interviews. Participants enjoyed being able to see the planets up close, and those in the VR condition more consistently focused on being immersed and able to move around within the simulation. Only a couple of participants in the PC conditions discussed the visualizations as something they enjoyed, while others talked about being able to "play around with it" (A204) or "seeing all of the details" (B303 and B304).

The audio groups also freely explained mappings during the follow-up interviews and discussed sounds they remembered, and mechanics of how they explored/used the sounds. All learners could recall facts they learned, with similar numbers reported by the PC conditions (Groups A & B) and the VR – no-audio condition (Group C). Group D had the highest number of recalled facts, with some being from the sound layer (e.g., D502 recalled "the noise, that, the sounds that – Saturn's rings, or any of the rings. Saturn's stood out b/c it was both visual and sound").

6.4 Application Domains

This study focused specifically on learner exploration of the solar system. Of course, these science concepts, at their base, are declarative knowledge; however, the basis of this activity was additionally grounded in the context of the NGSS and the Common Core. In both of these, both declarative knowledge, and deeper understanding and interpretation of models are key elements. Simulations, even ones which focus on presenting declarative knowledge, can be used as a platform for comparing information, key ideas, and helping learners develop understanding about larger systems.

Previous research has shown the application of VR in training for building or working with complex systems, particularly in manufacturing, and demonstrated its usefulness for helping students understand complex models, like chemical bonds or human anatomy. While the use of VR in this particular context (astronomy) may seem overly complex, this is of course, a simple model; this study focused on younger learners and some of the basic information they are expected to know. Ideas within this study (sound-enhancement or the multimedia effect, more broadly) could be integrated into the regular version of the Universe Sandbox (for example), which could be used to support exploration and inquiry for learners at different ages and knowledge levels. Space is a particularly relevant case for VR, as a fully-immersive environment can convey things like size and scale, which are not well represented through traditional print media (i.e., the typical small diagrams used to represent the planets in a condensed space). VR environments may also better convey comprehensive differences in visual scale (or other model components) compared to other digital media (i.e., PC simulations) (Jang et al., 2017).

In this study, the multimodal VR condition had particularly advantageous outcomes in supporting higher levels of intrinsic motivation, active cognitive engagement, physical engagement, and task mastery than the other conditions; even the multimodal PC condition had higher levels for some of these crucial factors compared to the visual-only VR condition. For other topics which may be declarative-focused, or may not lend themselves immediately to VR, the consideration of multimodal (at either a PC or VR level) could significantly influence the way the learners explore the content. Leveraging audio also provides another way to support data literacy for learners, through incorporating the vocabulary and asking them to critically discuss or reflect on the information gained from the system. It provides another means to scaffold discussion around, and even in cases where VR may not be the most relevant presentation means, the incorporation of audio may lead to further learning.

6.5 Limitations and Participant Considerations

Although the sample size for Study 3 was large enough to find differences between groups, a larger sample could improve statistical power and allow for group comparisons between all four conditions. This sample size was limited, in part due to the COVID-19 epidemic. For example, one group of 12 participants canceled as schools were understandably closed for health and safety precautions, drastically changing the ability to collect data in-person with students. These additional students may have provided enough power to complete Kruskal-Wallis tests (non-parametric equivalents to the one-way analysis of variance) to compare between all four groups instead of independent group differences which were completed for the VR & PC and audio and no-audio conditions. Even with the small sample size, there were significant differences between the VR and PC conditions (superficial cognitive engagement and task mastery) and between the audio and no-audio conditions (intrinsic motivation, active engagement, and task mastery). The chi-square analyses supported quantitative comparisons between the qualitatively coded data and gave additional insight into group differences which could not be evaluated through statistical means.

Another consideration, which while not necessarily a limitation in the study itself, is a limitation in how representative the recruited students are of a general student population. Many of these students may have been homeschooled and were recruited through CEISMC's First Lego League (FLL) homeschooled or community coaches. Recruitment for this study was wide and varied: at least five of the students were referred through a connection at Robotic Explorers (a STEM group in Roswell, GA); one student was referred from Sunshine Steam (a local STEAM summer camp); another student learned of the study from the STEAM Education Advancement Summer Camp; one participant came from a referral by the South Dekalb Improvement Association; two were from FLL teams at a local recreation center; and one learned of the study from a Facebook post. Three other students were recruited to participate from the greater Atlanta area private school.

It is difficult to say for certain how similar this simulation experience might have been across all of these groups, or how much the study results transfer from the enrolled participants to a broader learner audience. These learners may have been more motivated to participate in the study, or to participate in other types of educational activities due to their community or other STEM group connections. Would other students from the Atlanta Public Schools have had similar experiences, outcomes, and motivational drive? As differences were found between groups even in this smaller-scale study, it supports the usefulness in completing a larger-scale study with a more diverse learner group in the future.

6.6 Future Research

Future research could improve our understanding of the impact of non-speech audio on the Multimedia learning effect. Even with this small sample size, learners in both multimodal conditions reacted positively toward the audio-enhanced simulation. A larger evaluation might evaluate only the multimodal PC and VR conditions, instead and be able to make stronger claims about differences between the technology. This study found presence differences for the VR conditions and differences in other coded behavioral physical engagement for the participants. The audio conditions also resulted in higher motivation and active engagement. Research has previously shown the impact of embodiment and physical engagement for immersive virtual environments, and when combined with the higher cognitive engagement for the VR conditions shown through the descriptive and inferential data, a larger impact on supporting better opportunities for learning may be found.

6.7 Conclusion

This primary question studied here was: How well can multimodal Virtual Reality systems support learning and engagement compared to typical interactive simulations for science education? It evaluated learner experiences between visual-only and multimodal interactive simulations. It also evaluated learner experiences between PC and VR simulations. It compared multiple factors including engagement (cognitive, emotional, physical), usability of the system, motivation, and short and long-term learning opportunities (details remembered and accuracy scores). This study also compared potential impacting factors on a student's use of new technology for science learning experiences like science anxiety, technology use, participation in additional science and educational media, forecasted affect, and perception towards science.

This study found statistically significant differences between the VR and the PC conditions, with the VR condition supporting lower superficial cognitive engagement and higher task mastery than the PC condition. It also found statistically higher intrinsic motivation, active cognitive engagement, and task mastery for the multimodal (audio-enhanced) condition compared to the no-audio condition. There were few statistically significant observed differences for prior influencing factors, and prior knowledge differences were mitigated between groups for the learning outcomes in the post-accuracy solar system questionnaire after simulation use.

CHAPTER 7. DISCUSSION

Chapter 6 focused on the discussion from Study 3 specifically, in particular using a previously-designed space sonification as a secondary information presentation method within the audio-enhanced simulations. This chapter reflects on the entirety of this dissertation, and particularly relevant outcomes for the education, auditory display, and VR communities. These items include interaction's effect on sonification interpretation, sound literacy, when VR or audio is relevant (for learning tools), the multimedia principle, evaluating auditory user experience, and the construct engagement. For many of these items, Study 3 is most-related due to its encapsulation of multiple concepts from Studies 1 and 2, and will often be discussed first.

7.1 Interaction's Effect on Sonification Interpretation

Interaction may directly influence how learners use and interpret sonifications. Differences in the amount of control and the types of data being represented through the display could affect the learner's experience.

One difference between these three studies (Study 1, 2, and 3) and other studies of audio-enhanced educational tools is the amount of active control over the sounds themselves. For this dissertation, the sound designs were specifically chosen to highlight content teachers identified in the interviews and to leverage mappings based on prior work in sonification and auditory displays. This meant that the sounds were specifically designed to represent a numerical value, a concept, or categorical information, instead of being parameter-mapped sonifications (which would change the sound as the input values changed: e.g., a dynamic increase in mass would result in a dynamic pitch decrease). While the Study 3 participants were the ones directing their exploration, tool use, and interpretation of the sounds, they were not involved in the creation of those sounds, nor were they actively involved in uncovering/discovering the exact nature of the mappings related to the data relationships (due, in part, to the study goals and the design of the activity).

Other work with simulations and interactive learning tools has looked at different ways to add sound to enhance learning for accessibility purposes (Lahav et al., 2016; Levy & Lahav, 2012; T. L. Smith et al., 2017). Prior work with adding sonification to the PhET simulations has focused on leveraging this exploration of the sound by learners, particularly since these simulations are built to use implicit scaffolding to support immediate exploration and interaction from learners (Podolefsky et al., 2013). Sounds for the PhET simulations are designed a variety of ways, through expert teacher feedback, iterative sonic information design (Winters et al., 2019), and numerous formative and summative evaluations (Tomlinson, Batterman, et al., 2018; Tomlinson, Kaini, et al., 2019; Tomlinson, Walker, et al., 2019, 2020).

Evaluations for the sound designs integrated into the PhET simulations are more similar to Study 2 in this dissertation: they are focused directly on understanding interpretation, preference, and usability of the sounds. One key difference is that evaluations for PhET simulations need to focus on the ability of a learner to quickly explore changes in the sound mappings and to create a mental model of those representations. Often, learners can rely on these sounds with a visual (numerical, animation, or ad-hoc interaction) representation, but as the initial purpose of these sounds was to support non-visual access in combination with other speech feedback, their immediacy of interpretation is quite important (T. L. Smith & Moore, 2020; Tomlinson, Walker, et al., 2020).

For the PhET simulations, learners interact, change values, and observe changes to the model after that interaction. In this process, they build knowledge of cases and relationships that the sounds represent. Participants in Study 3 did not have this same type of experience; they were oriented to the Universe Sandbox, given a short introduction to the sound mappings, and then immediately began the activity. The goal was not to explore and uncover the sound mappings, but instead to rely on those sounds as a secondary means to explore the information about the planets and to make comparisons between those sound mappings. That is not to say that one type of interaction is necessarily better or worse. Someone could easily modify Study 3 to instead have learners work to build that type of knowledge during exploration; however, it was not the purpose of this dissertation, so the activity was thusly structured.

No matter the experience, the breadth of recent work exploring the integration of sound and other non-visual modalities (e.g., speech) into simulations has presented numerous cases supporting the benefit of sounds in such technology. Sounds could be integrated for additional representation methods, for accessible ad-hoc feedback, and even as a way to support student observation and data-collection for other classroom activities or experiments. The limits of sound are not restricted to particular cases, and while sound has been historically used for accessibility (L. M. Brown et al., 2002; L. M. Brown & Brewster, 2003; Davison, 2013; Kramer, 1994; Tomlinson, Batterman, et al., 2016; Walker & Nees, 2011), the results of these dissertation studies and other work with interactive simulations (Tomlinson, Walker, et al., 2019) should encourage the community to branch out. We are beyond the need to rely on typical ideas of sonification (i.e., solely for accessibility) and should integrate them more thoroughly into a variety of activities with varying levels of learner control, use, and even autonomy in sound design.

7.2 Sound Literacy

Generally, people have less experience with sound and talking critically about sound. For studies where sound represents a central modality, participants may struggle to discuss or feel confident discussing sounds. Building someone's knowledge of sound and sound literacy could directly influence how they interact with sound in everyday situations; it may also influence student learning experiences by providing additional means of exploration and context for students to situate individual, pair, and group learning activities.

Sound literacy, and general understanding of sound in classroom environments could come from a few different areas. For example, both the Georgia Standards (Georgia Department of Education, 2016) and the Next Generation Science Standards (National Research Council, 2013) require students to understand properties of sound waves. Other students may have the opportunity to learn about sound through a variety of music classes. However, exposure to these concepts does not guarantee learners will have a thorough understanding about sound or the terminology used to have discourse about it.

Even in our everyday environments, as much as people are generally exposed to (and rely on) sound, most do not spend time critically thinking about or analyzing sounds around them. The auditory system is amazing at parsing the sounds we hear, and helps us contextualize the world around us (Bregman, 1994). However, there is a general lack of critical analysis and discussion about these sounds. Culturally, we have focused on integrating this analysis into our visual environment; people can identify designs they like, phone app layouts which are easy to use or familiar, and have a basis in the language needed to discuss them. Unfortunately, this ubiquitous knowledge is not true of audio. Much of this knowledge is specific to fields like Foley sound design (Ament, 2014; Hug & Misdariis, 2011; Taylor, 2017), movie soundtrack design (Hillman & Pauletto, 2014), and video game design (Alves & Roque, 2010; Rogers, 2017; Tan et al., 2010). Understanding best practices for adding this knowledge into educational curricula is still a recent effort (Kemper, 2014). Addressing this expansion of knowledge and understanding of sound terminology, sound design, and critical analysis of audio in our environment is the goal of a recent Coursera Course (Tomlinson, Moore, et al., 2020).

Integrating sound literacy into the classroom has the potential to support many learners. Sound can help students with impairment actively explore information they may not otherwise have access to in a classroom environment (Tomlinson, Kaini, et al., 2019), and can be useful for a more diverse group of learners (Tomlinson, Batterman, et al., 2018). Sound can leverage different metaphors to support student engagement (Antle et al., 2008; Bakker et al., 2009). Purposefully-designed, thoughtful sounds can support more than this: interactive sounds and sonification can support data exploration (Grond & Hermann, 2014); independent collection of data in classroom experiments (Lahav et al., 2016); and additional visualization methods (L. M. Brown et al., 2002; Diaz-Merced et al., 2011; Ramloll et al., 2001, 2000). Educational environments (formal and informal) should seriously consider the benefits and the current difficulties related to sound literacy, as integration of these ideas into learning contexts would provide another means to support independent student learning.

7.3 When is VR or Audio Particularly Relevant?

Since the early days of VR, its potential for supporting immersive environments has excited researchers (Steuer, 1992), and this idea is still being explored today (Allcoat & von Mühlenen, 2018). Understanding when VR and audio are supportive of learning environments is an important consideration for researchers, teachers, and school districts.

This dissertation evaluated a few important application areas for educational technology; in particular, within Study 3, it focused on learner use of PC simulations, VR simulations, and auditory displays (Studies 1 and 2 evaluated these displays as well). In general, the PC simulations were successful at supporting opportunities for learning in Study 3. Learners in the two PC conditions did show an increase from their prior knowledge levels of the solar system concepts in the post-activity questionnaires. The VR simulation conditions were also successful at supporting opportunities for learning, and the difference in simulation media led to higher levels of physical and cognitive engagement. Learners in the audio conditions also had higher intrinsic motivation and higher cognitive engagement scores than the other groups. Evaluating differences in the between the two VR groups demonstrated an additional effect of the audio: the VR – audio group had higher physical engagement and relied less on the visual representations than the VR – no-audio group.

Studies 1 and 2 evaluated the design of the solar system sonification. That sonification was created initially as an informal learning experience for a planetarium. Planetarium shows typically rely on spoken description to scaffold an audience member's visual experience. Study 1 evaluated the audience's experience at a planetarium show including a combination of visuals, spoken description, and non-speech audio. Many audience members had positive reflections about the information, and described how it made them consider aspects of the solar system differently (Tomlinson et al., 2017). Study 2's focus was on a couple of things: evaluating the ability for sound to directly influence learner knowledge and supporting the creation of a scale to measure audio user experience. Study 2 found that learners *did* have an increase in accuracy from the pre- to post-test, and provided a basis for the creation of the BUZZ scale (Tomlinson, Noah, et al., 2018).

Most educational auditory displays are created with the purpose of supporting learners with vision impairment (Bonebright et al., 2001; Walker & Nees, 2011), and some additionally use audio as a means to further engage students (Paterson et al., 2010). While many in the auditory display community have long-discussed and presented research on the ability of sound to support a general audience (Kramer et al., 1999), the broader scientific community has not integrated this into their own work (Nees, 2018). The auditory display research community has also reported the potential for these displays to improve recognition and recall (Flowers, 2005), as well as engagement (Upson, 2001, 2002). The positive results of Studies 1 and 2 led to the final design of Study 3, as a more comprehensive evaluation directly comparing a visual-only and audio-enhanced learning tool had not been completed at this level previously.

While Study 3 found that the VR versions of the simulation had a positive impact on the learners, other factors (such as the financial cost associated with integrating these types of tools into the classroom) should be considered. The cost of professionally-available, commercial headsets has drastically reduced over the last few years, due to improvements in technology and their wider availability to a general audience (e.g., the Vive Cosmos is

$699[7]; the Oculus Quest is $499[8]). Many of them no longer require a hardwired connection to a computer (e.g., the Vive has a wireless adapter kit[9]); some require no additional hardware at all (e.g., the Oculus Go[10]). However, expecting school districts to use large portions of their budget for sophisticated VR headsets on the hope of using visual-only simulations may be unreasonable (Neelakantan, 2019). Additional barriers might include things like technical support from school staff and availability of materials for teacher curriculum planning. The positive impact of the combination of the VR *and* audio could tip the scales in favor of VR, as that combination led to the highest levels of physical engagement, cognitive engagement, and motivation. In cases where VR classroom integration is not available, Study 3 also confirms the positive effect of the PC – audio simulation on student learning opportunities, making it a reasonable solution in the meantime.

Additionally, the results of these three studies support the overall positive effect of audio on learning experiences. Audio should be integrated into a variety of materials, including informal learning experiences (such as museums or aquariums, which have some history of auditory display for accessibility purposes (Bruce & Walker, 2010; Jeon et al., 2012; Walker et al., 2006)), classrooms (Tomlinson, Batterman, et al., 2016), individual learning activities (Zhao et al., 2005), data analysis (Flowers et al., 2005), and even paired/group activities (Gaver et al., 1991). Audio has the potential for supporting higher levels of cognitive engagement, physical engagement, intrinsic motivation; it can also be

[7] https://www.vive.com/us/product/vive-cosmos/overview/
[8] https://www.oculus.com/quest/?locale=en_US
[9] https://www.vive.com/eu/accessory/cosmos-wireless-adapter-attachment-kit/
[10] https://www.oculus.com/go/

used as a primary means of supporting access for learning (Batterman & Walker, 2012), for fun (Batterman et al., 2013), and also for everyday tasks (Tomlinson, Schuett, et al., 2016). Even if VR is not widely-integrated into classroom learning experiences, audio and auditory displays could (and should) be added to support learners.

7.4 Multimedia Principle

Many educational tools have used the multimedia principle to engage students in innovative activities with the hope of creating circumstances which support better learning and problem-solving; however, there has been a distinct lack of evaluation for non-speech auditory displays under the principle.

The multimedia principle, first presented by Mayer (2002a, 2002b), has impacted the education community significantly over the last twenty years (for example, his Psychology of Learning and Motivation paper has over 11,700 citations[11]). One concern about integrating multimedia experiences into educational tools could be the impact on cognitive load (Low et al., 2010; Low & Sweller, 2014), and more particularly, a significant change to germane cognitive load, which is necessary for learning (Sweller, 2011).

Across all three studies, a multimedia environment was used to present the information. For Studies 1 and 2, three layers of information were presented: visual, non-speech audio, and speech; these three layers worked together to scaffold learner experience so they did not have to rely on just one media or the other. Participants in these studies did not complain about being overwhelmed by the amount of information, and even in the more

[11] https://scholar.google.com/citations?hl=en&user=o5doXYoAAAAJ&view_op=list_works

complex representations (Study 2's planetary perspective) could still easily interpret the information to correctly answer questions. Study 3 also found that participants had high accuracy for interpreting the sonifications, and at the same time, the audio conditions did not have any significant difference in overall usability.

This follows the different theoretical perspectives on human cognition for these multiple types of media. Dual Coding Theory (Clark & Paivio, 1991) has previously noted differences in verbal and non-verbal processing, which is in line with Wickens' Multiple Resource Theory (Wickens, 2008) and even Baddeley's separated working memory types based on sensory modality (Baddeley, 1992, 2010). Research building on the multimedia principle has focused on text (Dubois & Vial, 2001), narration (Harskamp et al., 2007), animations (M. Liu et al., 2008), and video (S. H. Liu et al., 2009). This dissertation addresses a major gap in multimedia research, by investigating non-speech audio, which (based on multiple cognition theories) should be able to support learning. Studies 1, 2, and 3 all provide evidence for the potential of non-speech audio to enhance learning, both with speech (Studies 1 and 2) and without (Study 3). The positive results of these studies should encourage researchers, designers, and educators to integrate non-speech audio into educational technologies as a new step forward for the multimedia principle.

7.5 Evaluating Auditory User Experience

Standardized evaluations make it easier to compare between multiple sound designs and select the best mappings. These evaluations also provide feedback about how learners may feel about the aesthetics, learnability, and give insight to their overall comprehension of the design mappings.

Measures of auditory user experience typically have been subject or application specific. Usually, these evaluations rely on individual (or a handful) of Likert or Likert-type questions. The problem is that without any standardization, like there has previously been within the HCI community (UMUX (Finstad, 2010), SUS (Lewis & Sauro, 2009), UMUX-Lite (Lewis et al., 2013), or SUPR-Q (Sauro, 2015)), it is difficult to know how successful the design of one auditory display is compared to another or whether iterations on a design clearly affect or change the user's perception.

Study 1 worked to develop an initial version of the audio user experience questions. Instead of following the typical scale design steps, each item in this evaluation was chosen based on similarly designed scales from visually glanceable displays (Matthews et al., 2007). As the planetarium audience would have only a brief amount of time listening to the sounds, leveraging these previous evaluations of ephemeral displays fit more closely than other typical usability evaluations. A few additional questions probed listeners to report their aesthetic responses, in order to understand the effect of aesthetics and enjoyment on interpretation and ease of use.

After completing Study 1, Study 2 focused on evaluating the solar system sonification through both an experiential perspective and its ability to support learning. It was important to validate whether or not listeners could listen through the sonification and interpret the underlying information in each display section. The first version of BUZZ, the audio user experience scale, was created for this study, to measure the effectiveness, usability, and overall aesthetic appeal of the display (Tomlinson, Noah, et al., 2018). While general usability or user experience scales could sufficiently measure a listener's experience, interactions with auditory displays may vary, and not always be easily comparable through something like UMUX (anecdotal feedback from study participants and other auditory display evaluations highlighted the need for a scale which better fits the audio context). BUZZ was created to address this need. Study 2 had two separate sets of auditory displays which were evaluated (the solar system perspective and the planetary perspective), and overall, two to three factors were found for the scale.

Study 2 presented preliminary work in the development of a measure of audio user experience, although a finalized version has not been released. This version of BUZZ was expanded to include more questions for Study 3 (which is more aligned with typical scale design studies). With the small sample of participants who used the audio-enhanced simulations in Study 3, an updated scale is not yet viable; however, additional work in the Sonification Lab is ongoing, and an updated scale should be available soon. The auditory display community has continued this dialog (regarding the development of a measure for audio user experience) through ongoing discussions at workshops and conferences (Tomlinson, Holthausen, et al., 2019). Future work in sonification, auditory display, and sound-enhanced multimedia should integrate the evaluation of auditory user experience into any evaluation.

7.6 What is Engagement, Really?

Engagement is a construct which may have significant influence on how students use and learn from an educational technology in classroom and non-classroom contexts. Understanding which types of engagement should be measured for particular applications and scenarios will help inform study design and potentially reduce study complexity.

At the beginning of this dissertation, the background and prior work reviewed a variety of previous definitions, measurements, scales, and use contexts for the term "engagement." As stated previously, engagement encompasses multiple levels or types: cognitive, intellectual, physical, emotional, academic, and social; engagement could also include or be influenced by factors like interest, effort, motivation, self-efficacy, time on task, or willingness to participate (Appleton et al., 2008; Fredricks et al., 2004; Pintrich & De Groot, 1990; Shernoff & Csikszentmihalyi, 2009).

Study 3 included a variety of measures from the fields of learning sciences, education, and HCI, with the goal of understanding *which* types of measures and tools are useful for evaluating different learners' experiences with interactive simulations. Some of the aforementioned engagement sub-categories were not relevant to this particular study context (e.g., social), and were not measured. Other types that have been previously studied, such as emotional engagement (Alexander et al., 2016; Fredricks et al., 2004; Yazzie-Mintz, 2007), were measured (e.g., using PANAS (Watson et al., 1988)), but no differences were found between groups at both the pre- and the post-activity time points. This lack of difference prompts an interesting question of the educational technology community: while some types of engagement have historically mattered in understanding student educational experiences, do they always matter, or are there cases where they can be omitted from future studies? Potentially, factors such as emotional engagement are important to students in a classroom context, where learners are situated within schools, in a larger community of other students, teachers, and administrators. Or, it may matter more in pair or group activities, where social engagement could affect learner experiences.

Measuring the components of engagement (and motivation, etc.) are crucial to actually evaluating and understanding differences in learner experiences. However, the individual factors which affect learners may differ between contexts. Educational technology researchers have traditionally measured many of these factors in classroom evaluations, school-wide surveys, and to a lesser extent, lab studies. The results of Study 3, particularly the lack of difference between some of these measurements (e.g., emotional engagement, some motivational components, science anxiety) for the different conditions might suggest a need to critically analyze which components are included in future evaluations. Certainly, including validated surveys and scales provided necessary measurements to explore between-group differences; however, they did take up a large portion of the study session, limiting the length of the exploration activity.

Future work in all educational technology evaluations should explore whether this is consistent across other contexts: should all components of engagement be measured, or are there pieces which can be removed due to the study constraints, technological factors, activity differences, or for other reasons? Building this knowledge may help reduce the amount of time and effort needed to recruit participants for longer studies, when a shorter, more succinct evaluation may be sufficient. This could also inform evaluations completed at a school or classroom level, making it easier to complete in situ educational technology studies.

7.7 Conclusion

This dissertation has explored the design and evaluation of a solar system sonification (Studies 1 and 2), a foil to the typical visual means used to convey astronomy information to an audience. The planetarium show and the lab evaluation were used to validate the mappings later included in Study 3.

It has also contributed new knowledge on how to measure multiple components of engagement, particularly for interactive simulations. It has compared different quantitative measures (validated scales from education research) and qualitative observational differences from video coding. Study 3 presented a structured lab study comparing learner experiences between four different types of simulations (PC – no-audio, PC – audio, VR – no-audio, and VR – audio). Overall, learners in the VR conditions had higher levels of cognitive and physical engagement than the PC conditions. Learners in the multimodal condition had higher scores for motivation and cognitive engagement than the no-audio condition.

Building from Studies 1 and 2, Study 3 provides additional evidence for the usefulness of non-speech audio (sonification) as an additional layer in the multimedia principle. It also demonstrates the usefulness of VR learning experiences, particularly when multimodal support is available. Finally, this thesis presented preliminary work for the development of a validated scale for auditory user experience (BUZZ). Overall, the results of these studies assert the need for greater inclusion of well-designed, validated auditory displays as a multimodal component of formal and informal learning experiences.

APPENDIX A. SCIENCE TEACHER INTERVIEWS

A.1 Semi-structured Question Guide

1. Can you tell me about your teaching background?
2. (If they need suggestions on where to start talking) Expanding on the survey?
3. What astronomy or space-related concepts are the students you teach tested on?
4. Can you give some examples?
5. Can you give a detailed example?
6. How do you initially introduce these topics?
7. Do you start with teaching Universe/Galaxies then move to Solar System or is it the opposite?
8. How do you structure building the knowledge of the solar system?
9. What aspect of the Solar System do you start with?
10. **Can you describe the structure you follow for how you teach the solar system?**
11. (Ask for as many details as possible – this is the main focus of the task analysis)
12. Ask for a run-through of the lesson plan and activities
13. Ask about the order of details taught, or the types of details chosen
14. What concepts confuse students the most?
15. How do they confuse the things you mentioned before?
16. What misconceptions do students have most commonly?
17. What strategies do you use to fix these misconceptions?
18. Is there any other information about teaching astronomy that you would like to share with us?

A.2 Demographics Questions

1. How many years have you taught?
2. Did you have an undergraduate major or minor in any of the following subjects?
    a. Biology or life science
    b. Chemistry
    c. Physics
    d. Geology or earth science
    e. Astronomy or space science
    f. Mathematics
    g. Elementary or secondary education
    h. Other(s): ______________________
3. Have you completed graduate coursework in any of the following subjects?

   a. Biology or life science
   b. Chemistry
   c. Physics
   d. Geology or earth science
   e. Astronomy or space science
   f. Mathematics
   g. Elementary or secondary education
   h. Other(s): ______________________________
4. Do you have any other background which helped to prepare you for teaching?
5. What grade(s) do you teach now?
6. What grade(s) have you taught?
7. What subject area(s) do you teach currently?
8. What subject area(s) have you taught previously?
9. How frequently do you update your teaching curriculum?
10. What general teaching methods do you use?
    a. Lectures
    b. Small-group work
    c. Group or individual projects
    d. Whole-class activities
    e. In class reading
    f. Using internet materials
    g. Lab activities
    h. Demonstrations
    i. Other(s): _______________________
11. What are the top 3 teaching methods you use?
12. What types of assessment do you give to your students?
    a. Homework
    b. Quizzes
    c. Tests
    d. Presentations
    e. Projects
    f. Other(s): _______________________
13. What are the top 3 assessment methods you use?
14. What types of teaching materials do you use in the classroom?
    a. Textbooks
    b. Other books
    c. Videos (DVDs or internet)
    d. Field trips
    e. Diagrams
    f. Whiteboard or blackboard
    g. Electronic smart board

h. Websites
i. Tablet applications (e.g. iPad)
j. Smartphone applications
k. Personally-developed materials
l. Adapted teaching materials
m. Physical models
n. Assistive technologies
o. Labs activities
p. Demonstrations
g. Other(s): ______________________

15. What are the top 3 teaching materials you use in the classroom?

APPENDIX B. MISCONCEPTION IDENTIFIER SURVEY

B.1 Misconception Survey Questions

Each question asks for an answer then a confidence rating, like:

(Question Text)?
Answer:
How confident are you, on a scale of 1 to 5, in that answer? ______
(1= I completely guessed/I am not confident at all,
2 = I am mostly not confident,
3 = I was about half-confident,
4 =I am mostly confident,
5=I did not guess at all/I am completely confident)

1. How many planets are in the solar system?
2. What are the names of the planets?
3. What is the order of the planets?
4. List the planets in order from smallest to largest diameter.
5. Are there any other natural orbiting bodies in the solar system between two consecutive planets? If so, what?
6. What are the main categories for planets based on composition?
7. What are some distinctive characteristic(s) about Mercury?
8. What are some distinctive characteristic(s) about Venus?
9. What are some distinctive characteristic(s) about Earth?
10. What are some distinctive characteristic(s) about Mars?
11. What are some distinctive characteristic(s) about Jupiter?
12. What are some distinctive characteristic(s) about Saturn?
13. What are some distinctive characteristic(s) about Uranus?
14. What are some distinctive characteristic(s) about Neptune?
15. Which planets have rings? List them, if any.
16. Which planets have moons? List them, if any.
17. What is Pluto categorized as?
18. What is the Sun?
19. What is gravity, and why is it important to the solar system? (Please write a complete sentence)
20. What causes seasons on Earth? (Please write a complete sentence)

21. What effect does the Sun have on Earth, if any? (Please write a complete sentence)

22. What effect does the Moon have on Earth, if any? (Please write a complete sentence)

23. What effect does the Earth have on the Moon, if any? (Please write a complete sentence)

24. What causes the phases of the moon? (Please write a complete sentence)

25. What is an orbit?

26. What is the difference between a planet's rotation and revolution?

27. On the back of this sheet of paper, draw to scale and as accurately as possible your own picture of the solar system.

28. Why did you draw the solar system the way you did? Or, did you have a reason for the representational style you chose?

B.2 Exit Survey Questions

1. Gender?
    a. Male ______
    b. Female ______
2. Age?
3. Major(s)?
4. Minor(s)?
5. What is the last class or grade in which you remember learning about space?
6. What informal learning activities do you enjoy?
    a. Museums
    b. Zoos
    c. Aquariums
    d. Planetariums
    e. Interactive Science Centers
    f. Other(s): ________________________________
7. Do you watch any of these television channels or shows?
    a. Cosmos
    b. NASA TV
    c. Discovery channel
    d. Science channel
    e. Billy Nye
    f. Other(s): ________________________________
8. Do you follow other science news?

a. On Facebook
b. On Twitter
c. Online news sources
d. Space magazines/websites
e. Other(s): ___________________________________

9. Have you ever attended a summer camp or a STEM (Science, Technology, Engineering, or Math) outreach day?
10. In the past, have you ever completed a research project for a science class on space?

APPENDIX C. SOLAR SYSTEM SONIFICATION

C.1 Recording

Recording available at: https://youtu.be/N_w7ST4rkW8?t=30s

C.2 Script Outline

- Short introduction and explanation of sonification [1 min.]
- Introduction & brief verbal overview of the planets (details selected where appropriate), to provide contextual situation and introduction to them in case the participant had no prior knowledge: location, size, rotation pattern, temperature, atmosphere, orbital period (revolution time), moons, rings. Each planet's mass sonification was played at the same time as the verbal intro. [5.5 min.]
    - Mercury
    - Venus
    - Earth
    - Mars
    - Asteroid Belt (used as delimiter between inner [planets before this item] and outer planets [planets after this item])
    - Jupiter
    - Saturn
    - Uranus
    - Neptune
- Sonification Model Part 1: includes comparisons to help introduce the scaling.
    - Mass: pitch-based [CITE previous work] [2.5 min.]
        - Mass of all the planets
        - Earth only
        - Mercury only
        - Venus only
        - Mars only
        - Jupiter only
        - Earth and Jupiter
        - Saturn and Jupiter
        - Uranus and Neptune
        - Earth, Uranus, Neptune
        - All planets again

    - Length of day: a ‘beating’ or ‘pulsing’ sound representing day and night for each planet [cite something]. [4 min.]
        - Earth only
        - Mercury only
        - Venus only
        - Mercury and Venus
        - Earth again
        - Mars
        - Mars and Earth
        - Jupiter
        - Saturn
        - Uranus
        - Neptune
        - All inner planets
        - All outer planets
        - All planets
    - Length of year: spatial audio of the planet’s location moving around the listener (panning through headphones). [2.5 min.]
        - Mercury
        - Venus
        - Earth
        - Mars
        - All inner planets
        - Jupiter (new reference point)
        - Saturn
        - Uranus
        - Neptune
- Alternate distance from the sun mapping: pitch-shifting a rocket-ship’s speed to represent increasing amount of distance. [3 min.]
- Sonification Model Part 2: includes details specific to each planet, such as presence of moons, rings, temperature, composition (terrestrial vs. gas) [8 min.]
    - Mercury: mean temperature, composition, gravitational strength
    - Venus: mean temperature and composition
    - Earth: moon, gravitational strength
    - Venus: gravitational strength
    - Earth: composition and temperature
    - Mars: moons, gravitational strength, mean temperature, composition
    - Jupiter: gravitational strength, rings, moons, composition, mean temperature

- Saturn: rings, moons, gravitational strength, composition, mean temperature
- Uranus: rings, moons
- Earth, Saturn, Neptune gravitational strength
- Neptune: rings, moons
- Uranus composition and mean temperature
- Neptune composition and mean temperature
- Temperature ranges for all planets presented serially

C.3 Planetarium Survey Questions

Instructions: Please circle the number that best represents your thoughts!

Part 1: Solar System View [information like size, length of day, rotation direction, length of year]

1. How helpful were the sounds?

Very unhelpful	Unhelpful	Somewhat unhelpful	Somewhat helpful	Helpful	Very helpful
1	2	3	4	5	6

2. How interesting were the sounds?

Very uninteresting	Uninteresting	Somewhat uninteresting	Somewhat interesting	Interesting	Very interesting
1	2	3	4	5	6

3. How pleasant were the sounds?

Very unpleasant	Unpleasant	Somewhat unpleasant	Somewhat pleasant	Pleasant	Very pleasant
1	2	3	4	5	6

4. How easy was it to understand the sounds?

Very difficult	Difficult	Somewhat difficult	Somewhat easy	Easy	Very easy
1	2	3	4	5	6

5. How relatable were the sounds to their ideas?

Very unrelatable	Unrelatable	Somewhat unrelatable	Somewhat relatable	Relatable	Very relatable

1	2	3	4	5	6

Part 2: Planetary View [information like the moons, rings, temperature, planet type, and gravity]

1. How helpful were the sounds?

Very unhelpful	Unhelpful	Somewhat unhelpful	Somewhat helpful	Helpful	Very helpful
1	2	3	4	5	6

2. How interesting were the sounds?

Very uninteresting	Uninteresting	Somewhat uninteresting	Somewhat interesting	Interesting	Very interesting
1	2	3	4	5	6

3. How pleasant were the sounds?

Very unpleasant	Unpleasant	Somewhat unpleasant	Somewhat pleasant	Pleasant	Very pleasant
1	2	3	4	5	6

4. How easy was it to understand the sounds?

Very difficult	Difficult	Somewhat difficult	Somewhat easy	Easy	Very easy
1	2	3	4	5	6

5. How relatable were the sounds to their ideas?

Very unrelatable	Unrelatable	Somewhat unrelatable	Somewhat relatable	Relatable	Very relatable
1	2	3	4	5	6

Part 3: Overall Composition [the last part of the show]

1. How helpful were the sounds?

Very unhelpful	Unhelpful	Somewhat unhelpful	Somewhat helpful	Helpful	Very helpful
1	2	3	4	5	6

2. How interesting were the sounds?

Very uninteresting	Uninteresting	Somewhat uninteresting	Somewhat interesting	Interesting	Very interesting
1	2	3	4	5	6

3. How pleasant were the sounds?

Very unpleasant	Unpleasant	Somewhat unpleasant	Somewhat pleasant	Pleasant	Very pleasant
1	2	3	4	5	6

4. How easy was it to understand the sounds?

Very difficult	Difficult	Somewhat difficult	Somewhat easy	Easy	Very easy
1	2	3	4	5	6

5. How relatable were the sounds to their ideas?

Very unrelatable	Unrelatable	Somewhat unrelatable	Somewhat relatable	Relatable	Very relatable
1	2	3	4	5	6

Free response:

1. Was there anything you really liked or disliked?
2. Did you have a favorite sound or set of sounds?
3. Based on what you know about the planets, do you think they were correctly represented through sound?
4. How did listening to the planets make you feel?
5. Did your understanding about the solar system change? If yes, how?
6. Did this help you appreciate more about our solar system?
7. Was there something you didn’t know before that you learned tonight?
8. Age: ____
9. Are you a student? ___ No ___Yes
    a. If yes, what is your grade level? ___
10. If you feel comfortable sharing, what school do you go to? _______

APPENDIX D. LISTENING ACTIVITY MATERIALS

D.1 Pre/Post Test

Start of Block: Block 1

Q1a List the names of the planets in order from closest to farthest from the sun.

____

Q1b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

End of Block: Block 1

Start of Block: Block 2

Q2a Rank the planets in order from smallest to largest diameter.

______ Earth
______ Jupiter
______ Mars
______ Mercury
______ Neptune
______ Saturn
______ Uranus
______ Venus

Q2b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q3a What are the two main categories for planets based on composition?

____

Q3b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q4a Which planets have rings?

- ☐ Earth
- ☐ Jupiter
- ☐ Mars
- ☐ Mercury
- ☐ Neptune
- ☐ Saturn
- ☐ Uranus
- ☐ Venus

Q4b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q5a What causes the seasons on Earth?

Axis tilt and the varying altitude of the Sun
Distance from the Sun
Global warming
Rotation speed

Q5b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q6a What is the difference between a planet's rotation and revolution?

Revolution	Rotation
______ When a planet or moon turns all the way around or spins on its axis one time.	______ When a planet or moon turns all the way around or spins on its axis one time.
______ When a planet or moon travels once around an object.	______ When a planet or moon travels once around an object.

Q6b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q7a What causes the phases of the moon?

The day of the month
The position of the Earth in its orbit around the sun
The angle we see the sunlit side of the moon as it revolves around the Earth
The Earth is blocking different amounts of light from the sun

Q7b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q8a What is gravity, and why is it important to the Solar System?

____

Q8b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q9a What are some distinctive characteristics about Venus?

- ☐ Closest in size to Earth
- ☐ No atmosphere
- ☐ No magnetic field
- ☐ Opposite rotation (spins backwards)
- ☐ Closest planet to the Sun
- ☐ Hottest planet
- ☐ Gas planet
- ☐ Has faint rings

Q9b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

Q10a What are some distinctive characteristics about Uranus?

- ☐ Farthest planet from the Sun
- ☐ 27 moons
- ☐ Rotates horizontally
- ☐ Great dark spot
- ☐ Has a solid core
- ☐ Ice giant
- ☐ Shortest day
- ☐ Irregular magnetic field

Q10b On a scale of not confident (1) to completely confident (5), how confident are you in that answer?

1 = I completely guessed/I am not confident at all
2 = I am mostly not confident
3 = I am about half-confident
4 = I am mostly confident
5 = I did not guess at all/I am completely confident

End of Block: Block 2

D.2 Listening Activity Questionnaire

Part 1 Questions: Solar System View [information like size, length of day, rotation direction, length of year]

Instructions Part 1:
Please open the link at the bottom of this page in a separate tab. Listen up until around 16:30 and then pause the video & answer the questions on the next page. https://youtu.be/N_w7ST4rkW8?t=30s

Scale Ratings: BUZZ
For the sounds in the previous section representing size of the planet, length of day, rotation direction, and length of year, please rate how you agree or disagree with the following statements.

(Ratings: Strongly Disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly Agree)

1. The sounds were helpful.
2. The sounds were interesting.
3. The sounds were pleasant.
4. The sounds were easy to understand.
5. The sounds were relatable to their ideas.
6. It's easy to match these sounds to their meanings.
7. It's difficult to understand how the sounds changed from one variable to the next, or one planet to the next.
8. Please select "Somewhat disagree."
9. It's fun to listen to these sounds.
10. It's boring to listen to these sounds.
11. It was confusing to listen to these sounds.
12. It was easy to understand what each of the sounds represented.

Scale Ratings: UMUX
Thinking about the sounds you just listened to for the first section, please rate how much you think the sounds could help you compare one planet to another (for size of the planet, length of day, rotation direction, and length of year).

(Ratings: Strongly Disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly Agree)

1. These sounds' capabilities meet my requirements.
2. Using these sounds is a frustrating experience.
3. These sounds are easy to use.
4. I have to spend too much time correcting things with these sounds.

Free Response:
Any comments? (i.e., are there any other sounds you think would better represent size of the planet, length of day, rotation direction, and length of year)

Multiple Choice:

1. Which planet has the longest day? (Mercury, **Venus**, Earth, Mars, Jupiter, Saturn, Uranus, Neptune) [select one]

2. Which planets have the shortest days? (Mercury, Venus, Earth, Mars, **Jupiter**, **Saturn**, Uranus, Neptune) [select multiple]
3. How does Uranus rotate? (clockwise, counterclockwise, **rolling on its side**) [select one]
4. Which planet is most like Earth in size? (Mercury, **Venus**, Earth, Mars, Jupiter, Saturn, Uranus, Neptune) [select one]

Free Response:

1. List one detail from the first section that you did not know before.
2. List one sound from the previous section that you really liked.
3. List one sound from the previous section that you disliked.

Part 2: Planetary View [information like the moons, rings, temperature, planet type, and gravity]

Instructions Part 2:
Instructions: restart the video (around 19:10 - or open the video here https://youtu.be/N_w7ST4rkW8?t=19m10s) and listen to the second half of the video. Once the second section is over (around 29:50) please pause the video and answer these questions.

Scale Ratings: BUZZ
For the sounds in the previous section representing moons, rings, temperature range, gravitational strength, and type of planet, please rate how you agree or disagree with the following statements.

(Ratings: Strongly Disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly Agree)

1. The sounds were helpful.
2. The sounds were interesting.
3. The sounds were pleasant.
4. The sounds were easy to understand.
5. The sounds were relatable to their ideas.
6. It's easy to match these sounds to their meanings.
7. It's difficult to understand how the sounds changed from one variable to the next, or one planet to the next.
8. Please select "Somewhat disagree."
9. It's fun to listen to these sounds.
10. It's boring to listen to these sounds.
11. It was confusing to listen to these sounds.
12. It was easy to understand what each of the sounds represented.

Scale Ratings: UMUX

Thinking about the sounds you just listened to for the first section, please rate how much you think the sounds could help you compare one planet to another (or compare moons, rings, temperature range, gravitational strength, and type of planet).

(Ratings: Strongly Disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly Agree)

1. These sounds' capabilities meet my requirements.
2. Using these sounds is a frustrating experience.
3. These sounds are easy to use.
4. I have to spend too much time correcting things with these sounds.

Multiple Choice:

1. About how many times further out is Neptune from the Sun than Mercury? (We know it didn't directly cover this, just make your best guess) (39, 65, **78**, 91 times) [select one]
2. About how many times larger is gravity on Jupiter than Earth? (1.5, 2.0, **2.5**, 3.0) [select one]
3. What are the coldest planets in the Solar System? (Mercury, Venus, Earth, Mars, Jupiter, Saturn, **Uranus**, **Neptune**) [select two]
4. What are the hottest planets in the solar system? (**Mercury**, **Venus**, Earth, Mars, Jupiter, Saturn, Uranus, Neptune) [select one]
5. How many moons does Mars have? (0, 1, **2**, 3) [select one]

Free Response:

1. List one detail from the previous section that you did not know before.
2. List one sound from the previous section that you really liked.
3. List one sound from the previous section that you disliked.
4. Did this help you appreciate more about our solar system?
5. How did listening to the planets make you feel and why?

D.3 Demographics

The demographics survey for this study is the same as the one in B.2 Exit Survey questions.

D.4 BUZZ: Audio User Experience Questionnaire

Full question set:

1. The sounds were helpful.

2. The sounds were interesting.
3. The sounds were pleasant.
4. The sounds were easy to understand.
5. The sounds were relatable to their ideas.
6. It was easy to match these sounds to their meanings.
7. It was difficult to understand how the sounds changed from one variable to the next.
8. It was fun to listen to these sounds.
9. It was boring to listen to these sounds.
10. It was confusing to listen to these sounds.
11. It was easy to understand what each of the sounds represented.

APPENDIX E. PRE-LAB ACTIVITY MATERIALS

E.1 Adapted Technology Experience Profile

Please indicate how often you have used any of the technologies listed below in the last year. This could be at home, school, or anywhere else. Select the choice that best fits your use of each technology.

	I Don't Know What It Is (1)	Not At All (2)	Once a Month (3)	Once a Week (4)	Everyday (5)
1. Computer	○	○	○	○	○
2. Camera	○	○	○	○	○
3. Music Player (e.g., iPod, mp3 player)	○	○	○	○	○
4. Smart board	○	○	○	○	○
5. LCD projector	○	○	○	○	○
6. Printer	○	○	○	○	○

7. Robots (e.g., LEGO Mindstorms, Cubo, robot dog)	○	○	○	○	○
8. Student Response Systems (e.g., classroom clickers)	○	○	○	○	○
9. Tablet (e.g., iPad, 2-in-1, Touchpad)	○	○	○	○	○
10. Webcam	○	○	○	○	○
11. Smartphone (e.g., iPhone, Android)	○	○	○	○	○
12. Smartwatch	○	○	○	○	○
13. Video games (e.g., PlayStation, XBOX, Nintendo Switch)	○	○	○	○	○
14. Hoverboard	○	○	○	○	○
15. Computer Games	○	○	○	○	○
16. Electronic Book Reader (e.g., Kindle, Nook)	○	○	○	○	○

17. Social Networking (e.g., Facebook, Instagram)	○	○	○	○	○
18. Augmented Reality Headset (e.g., HoloLens)	○	○	○	○	○
19. Virtual Reality Headset (e.g., Vive, Oculus)	○	○	○	○	○

Did you use any of the technologies previously, but don't use them any longer? If so, list which ones: ____

E.2 Modified Attitudes Toward Science Instrument (mATSI)

The following statements are about the study of science. Please read each statement carefully. Use the following scale to show how much you agree or disagree with each statement. Answer these questions thinking about science or school in general.

	Strongly Disagree (1)	Disagree (2)	Undecided (3)	Agree (4)	Strongly Agree (5)
1. Science is useful in helping to solve the problems of everyday life.	○	○	○	○	○
2. Science is something that I enjoy very much.	○	○	○	○	○
3. I would like to do some extra or un-assigned reading in science.	○	○	○	○	○
4. Science is easy for me.	○	○	○	○	○
5. When I hear the word science, I have a feeling of dislike.	○	○	○	○	○
6. Most people should study some science.	○	○	○	○	○

7. Sometimes I read ahead in our science book.	○	○	○	○	○
8. Science is helpful in understanding today’s world.	○	○	○	○	○
9. I usually understand what we are talking about in science.	○	○	○	○	○
10. Science teachers make science interesting.	○	○	○	○	○
11. No matter how hard I try, I cannot understand science.	○	○	○	○	○
12. I feel tense when someone talks to me about science.	○	○	○	○	○
13. Science teachers present material in a clear way.	○	○	○	○	○
14. I often think, “I cannot do this,” when a science assignment seems hard.	○	○	○	○	○

15. Science is of great importance to a country's development.	○	○	○	○	○
16. It is important to know science in order to get a good job.	○	○	○	○	○
17. I like the challenge of science assignments.	○	○	○	○	○
18. It makes me nervous to even think about doing science.	○	○	○	○	○
19. It scares me to have to take a science class.	○	○	○	○	○
20. Science teachers are willing to give us individual help.	○	○	○	○	○
21. It is important to me to understand the work I do in science class.	○	○	○	○	○
22. I have a good feeling toward science.	○	○	○	○	○

23. Science is one of my favorite subjects.	○	○	○	○	○
24. I have a real desire to learn science.	○	○	○	○	○
25. I do not do very well in science.	○	○	○	○	○

E.3 Positive and Negative Affect Schedule (PANAS)

E.3.1 Group A Prompt

This scale consists of a number of words that describe different feelings and emotions.

Imagine you are going to use an interactive simulation on a computer. This simulation will be used to display information about our solar system.

Thinking about what this experience may be like, to what degree do you think you would experience each of the following feelings or emotions while using the interactive computer simulation.

E.3.2 Group B Prompt

This scale consists of a number of words that describe different feelings and emotions.

Imagine you are going to use an interactive simulation with sounds on a computer. This simulation will be used to display information about our solar system.

Thinking about what this experience may be like, to what degree do you think you would experience each of the following feelings or emotions while using the interactive computer simulation.

E.3.3 Group C Prompt

This scale consists of a number of words that describe different feelings and emotions.

Imagine you are going to use an interactive simulation on a virtual reality headset. This simulation will be used to display information about our solar system.

Thinking about what this experience may be like, to what degree do you think you would experience each of the following feelings or emotions while using the interactive virtual reality simulation.

E.3.4 Group D Prompt

This scale consists of a number of words that describe different feelings and emotions.

Imagine you are going to use an interactive simulation with sounds on a virtual reality headset. This simulation will be used to display information about our solar system.

Thinking about what this experience may be like, to what degree do you think you would experience each of the following feelings or emotions while using the interactive virtual reality simulation with sounds.

E.3.5 PANAS questions

	Very Slightly or Not at All (1)	A Little (2)	Moderately (3)	Quite a Bit (4)	Extremely (5)
Interested (curious or want to know more)	○	○	○	○	○
Distressed (anxious or upset)	○	○	○	○	○
Excited (eager or want to do more)	○	○	○	○	○
Upset (unhappy or worried)	○	○	○	○	○
Strong (powerful or can do things well)	○	○	○	○	○
Guilty (wrong or feel sorry)	○	○	○	○	○

Scared (afraid or nervous)	○	○	○	○	○
Hostile (unfriendly or mean)	○	○	○	○	○
Enthusiastic (feeling joyful or pleased)	○	○	○	○	○
Proud (pleased or happy with yourself)	○	○	○	○	○
Irritable (grumpy or upset)	○	○	○	○	○
Alert (awake or quick to understand)	○	○	○	○	○
Ashamed (guilty or sorry)	○	○	○	○	○
Inspired (encouraged or motivated)	○	○	○	○	○
Nervous (jumpy or tense)	○	○	○	○	○
Determined (stubborn or have a strong desire)	○	○	○	○	○

Attentive (alert or thoughtful)	○	○	○	○	○
Jittery (nervous or jumpy)	○	○	○	○	○
Active (full of energy or lively)	○	○	○	○	○
Afraid (scared or terrified)	○	○	○	○	○

E.4 Pre-Activity Solar System Questions

Next, you'll answer some questions about the solar system. Do your best, and try to answer all of the questions, even if you have to guess. If you aren't sure, it's ok to write "I don't know"

1. List the names of the planets in order from closest to farthest from the sun.
2. What are the two main categories for planets, based on composition?
3. What are the two largest planets?
4. What are the two smallest planets?
5. Which planets have rings?
6. Which planets have moons?
7. What are the two coldest planets?
8. What is the hottest planet?
9. List the names of the planets in order from largest to smallest (based on mass).
10. Name two planets that have similar surface gravitational strength.
11. Which planets have both moons and rings?
12. Which planet(s) have a day longer than Earth's?
13. Which planet(s) have a day shorter than Earth's?
14. Which planet has the most rings?
15. Which planet has the most moons?
16. What's different about the density or composition of planets with rings and those without rings in our Solar System?

E.5 Sim Sickness Questionnaire (SSQ)

Please indicate the extent to which you are experiencing the following symptoms:

	None (1)	Slight (2)	Moderate (3)	Severe (4)
General discomfort	○	○	○	○
Fatigue	○	○	○	○
Headache	○	○	○	○
Eye strain	○	○	○	○
Difficulty focusing	○	○	○	○
Increased salivation	○	○	○	○
Sweating	○	○	○	○
Nausea	○	○	○	○

Difficulty concentrating	○	○	○	○
"Fullness" of the head	○	○	○	○
Blurred vision	○	○	○	○
Dizziness (eyes open)	○	○	○	○
Dizziness (eyes closed)	○	○	○	○
Vertigo	○	○	○	○
Stomach awareness	○	○	○	○
Burping	○	○	○	○

APPENDIX F. LAB ACTIVITY MATERIALS

These prompts were used to encourage exploration and play during simulation use. The set learners use depends on their scores on the pre-test. If they answered questions relevant to the first prompts correctly on the pre-test, the experimenter would instead start the activity with Prompt Set 3 or 4.

F.1 Prompt Set 1 & 2

1. Explore and discover the order of the planets.
2. Explore and discover which planets have the 2 largest masses.
3. Explore and discover which planets have the 2 smallest masses.
4. Explore and discover which planets have rings.
5. Explore and discover which planets have moons.
6. Explore and discover what the 2 coldest planets are.
7. Explore and discover what the 1 hottest planet is.
8.

F.2 Prompt Set 3

1. Explore and discover which planet has the 3rd largest mass.
2. Explore and discover a planet with the most similar gravitational strength to Mercury.
3. Explore and discover a planet with the most similar gravitational strength to Venus.
4. Explore and discover which planets have both rings and moons.

F.3 Prompt Set 4

1. Explore and discover which planet has the most rings.
2. Explore and discover a planet with a day longer than Earth's.
3. Explore and discover a planet with a day shorter than Earth's.
4. Part 1: Compare 2 gas giants. What's similar about their densities?
   Part 2: Now look at Earth – is it similar or different than them?

APPENDIX G. POST ACTIVITY MATERIALS

G.1 Post-Activity Solar System Questions

Next, you'll answer some questions about the solar system. Do your best, and try to answer all of the questions, even if you have to guess. If you aren't sure, it's ok to write "I don't know"

1. List the names of the planets in order from closest to farthest from the sun.
2. What are the two main categories for planets, based on composition?
3. What are the two largest planets?
4. What are the two smallest planets?
5. Which planets have rings?
6. Which planets have moons?
7. What are the two coldest planets?
8. What is the hottest planet?
9. List the names of the planets in order from largest to smallest (based on mass).
10. Name two planets that have similar surface gravitational strength.
11. Which planets have both moons and rings?
12. Which planet(s) have a day longer than Earth's?
13. Which planet(s) have a day shorter than Earth's?
14. Which planet has the most rings?
15. Which planet has the most moons?
16. What's different about the density or composition of planets with rings and those without rings in our Solar System?

G.2 Positive and Negative Affect Schedule (PANAS)

G.2.1 Group A Prompt

Please respond to the following items to tell how much you feel each emotion right now based on using the interactive computer simulation.

G.2.2 Group B Prompt

Please respond to the following items to tell how much you feel each emotion right now based on using the interactive computer simulation with sounds.

G.2.3 Group C Prompt

Please respond to the following items to tell how much you feel each emotion right now based on using the interactive computer simulation on a virtual reality headset.

G.2.4 Group D Prompt

Please respond to the following items to tell how much you feel each emotion right now based on using the interactive computer simulation with sounds on a virtual reality headset.

G.2.5 PANAS questions

Items from this questionnaire are the same as those in the section E.3.5.

G.3 Slater-Usoh-Steed Presence Scale (SUS)

For these questions, "virtual environment" means The Universe Sandbox.

1. Please rate your sense of being in the virtual environment, on the following scale from 1 to 7, where 7 represents your normal experience of being in a place.

I had a sense of “being there” in the virtual environment.

- ○ 1 Not At All
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7 Very Much

2. To what extent were there times during the experience when the virtual environment was the reality for you?

There were times during the experience when the virtual environment was the reality for me...

- ○ 1 At No Time
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7 Almost All The Time

3. When you think back about your experience, do you think of the virtual environment more as images that you saw, or more as somewhere that you visited?

The virtual environment seems to me to be more like...

- ○ 1 Images That I Saw
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7 Somewhere That I Visited

4. During the time of the experience, which was strongest on the whole, your sense of being in the virtual environment, or of being elsewhere?

I had a stronger sense of...

- ○ 1 Being Elsewhere
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7 Being In a Virtual Environment

5. Consider your memory of being in the virtual environment. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual memory of the virtual environment, whether that memory is in color, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements.

I think of the virtual environment as a place in a way similar to other places that I've been today...

- 1 Not At All
- 2
- 3
- 4
- 5
- 6
- 7 Very Much So

6. During the time of the experience, did you often think to yourself that you were actually in the virtual environment? During the experience I often thought that I was really standing in the virtual environment...

- 1 Not Very Often
- 2
- 3
- 4
- 5
- 6
- 7 Very Much So

G.4 Usability Metric for User Experience (UMUX)

Thinking about the simulation, Universe Sandbox, that you just used, please rate how much you agree or disagree with the following statements.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
The Universe Sandbox's capabilities meet my requirements.							
Using the Universe Sandbox is a frustrating experience.							
The Universe Sandbox is easy to use.							
I have to spend too much time correcting things with the Universe Sandbox.							

G.5 Audio User Experience Scale (BUZZ)

Thinking about the set of sounds you just listened to, please rate how much you agree or disagree with the following statements. [Answered by the TWO audio conditions only]

	Strongly Disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
1. The sounds were helpful.	○	○	○	○	○	○	○
2. The sounds were interesting.	○	○	○	○	○	○	○
3. The sounds were pleasant.	○	○	○	○	○	○	○
4. The sounds were easy to understand.	○	○	○	○	○	○	○
5. The sounds were relatable to their ideas.	○	○	○	○	○	○	○
6. It was easy to match these sounds to their meanings.	○	○	○	○	○	○	○

7. It was difficult to understand how the sounds changed from one variable to the next.	○	○	○	○	○	○	○
8. It was fun to listen to these sounds.	○	○	○	○	○	○	○
9. It was boring to listen to these sounds.	○	○	○	○	○	○	○
10. It was confusing to listen to these sounds.	○	○	○	○	○	○	○
11. It was easy to understand what each of the sounds represented.	○	○	○	○	○	○	○
12. It was difficult to hear the changes in the sounds.	○	○	○	○	○	○	○
13. It was easy to hear the changes in the sounds.	○	○	○	○	○	○	○

14. The sounds did not match the ideas that they were intended to represent based on their application or context.	○	○	○	○	○	○	○
15. The sounds matched the ideas that they were intended to represent based on their application or context.	○	○	○	○	○	○	○
16. It was difficult to hear the differences between the sounds.	○	○	○	○	○	○	○
17. It was easy to hear the differences between the sounds.	○	○	○	○	○	○	○
18. It was difficult to compare the characteristics of each sound.	○	○	○	○	○	○	○
19. It was easy to compare the characteristics of each sound.	○	○	○	○	○	○	○

20. It was difficult to match these sounds to their meanings.	○	○	○	○	○	○	○
21. It was easy to match these sounds to their meanings.	○	○	○	○	○	○	○
22. It was easy to determine the meaning of changes in the sounds over time.	○	○	○	○	○	○	○
23. It was difficult to determine the meaning of the changes in the sounds over time.	○	○	○	○	○	○	○
24. It was easy to determine the meaning of the changes between the sounds.	○	○	○	○	○	○	○
25. It was difficult to determine the meaning of the changes between the sounds.	○	○	○	○	○	○	○

26. It would take a long time to be able to complete tasks using these sounds.	○	○	○	○	○	○	○
27. It would take a short time to be able to complete tasks using these sounds.	○	○	○	○	○	○	○
28. It would take a long time to be able to understand what changed in these sounds.	○	○	○	○	○	○	○
29. It would take a short time to be able to understand what changed in these sounds.	○	○	○	○	○	○	○
30. It would take a long time to learn the meaning of these sounds.	○	○	○	○	○	○	○
31. It would take a short time to learn the meaning of these sounds.	○	○	○	○	○	○	○

32. The changes in the sounds reflected the changes in the information.	○	○	○	○	○	○	○
33. The changes in the sounds did not reflect the changes in the information.	○	○	○	○	○	○	○
34. It was easy to interpret the meaning of one sound compared to another.	○	○	○	○	○	○	○
35. It was difficult to interpret the meaning of one sound compared to another.	○	○	○	○	○	○	○
36. The sounds were not helpful.	○	○	○	○	○	○	○
37. The sounds were not interesting.	○	○	○	○	○	○	○
38. The sounds were not pleasant.	○	○	○	○	○	○	○

39. The sounds were not easy to understand.	○	○	○	○	○	○	○
40. The sounds were not relatable to their ideas.	○	○	○	○	○	○	○

G.6 Motivated Strategies for Learning Questionnaire (MSLQ)

The following questions ask about your motivation for and attitudes about school in general and this learning activity. Remember there are no right or wrong answers, just answer as accurately as possible. Using the scale below to answer the questions.

If you think the statement is very true of you, choose 7; if a statement is not at all true of you, choose 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes you.

	1 (Not at all True of Me)	2	3	4	5	6	7 (Very True of Me)
1. In a learning activity like this, I prefer course material that really challenges me so I can learn new things.	○						○
2. If I study in appropriate ways, then I will be able to learn the material in this learning activity.	○						○
3. When I take a test I think about how poorly I am doing compared to other students.	○						○
4. I think I will be able to use what I learn in this learning activity in other courses.	○						○
5. I believe I will receive an excellent score in this learning activity.	○						○
6. I’m certain I can understand the most difficult material presented in the content for this learning activity.	○						○

7. Getting a good score in this learning activity is the most satisfying things for me right now.	○						○
8. When I take a test I think about items on the other parts of the test I can't answer.	○						○
9. It is my own fault if I don't learn the material in this learning activity.	○						○
10. It is important for me to learn the material in this learning activity.	○						○
11. The most important thing for me right now is improving my overall score so my main concern in this class is getting a good score.	○						○
12. I'm confident I can learn the basic concepts taught in this learning activity.	○						○
13. If I can, I want to get better score in this learning activity than most of the other students.	○						○
14. When I take tests I think of the consequences of failing.	○						○
15. I'm confident I can understand the most complex material presented by the simulation in this learning activity.	○						○
16. In a learning activity like this, I prefer course material that arouses my curiosity, even if it is difficult to learn.	○						○
17. I am very interested in the content area of this learning activity.	○						○

18. If I try hard enough, then I will understand the learning activity material.	○						○
19. I have an uneasy, upset feeling when I take an exam.	○						○
20. I'm confident I can do an excellent job on the assignments and tests in this learning activity.	○						○
21. I expect to do well in this learning activity.	○						○
22. The most satisfying thing for me in this learning activity is trying to understand the content as thoroughly as possible.	○						○
23. I think the material in this learning activity is useful for me to learn.	○						○
24. When I have the opportunity in this learning activity, I choose activities that I can learn from even if they don't guarantee a good score.	○						○
25. If I don't understand the learning activity material, it is because I didn't try hard enough.	○						○
26. I like the subject matter of this learning activity.	○						○
27. Understanding the subject matter of this learning activity is very important to me.	○						○
28. I feel my heart beating fast when I take an exam.	○						○

29. I’m certain I can master the skills being taught in this learning activity.	○						○
30. I want to do well in this learning activity because it’s important to show my ability.	○						○
31. Considering the difficulty of this activity, the program, and my skills, I think did well in this learning activity.	○						○

G.7 Science Activity Questionnaire (SAQ)

Students have a lot of different thoughts and feelings while they are doing science activities. We want to know how true each of these things below was for you. Here "work" means the science activity you completed

Remember there are no right and wrong answers. Select the answer that best describes your feelings.

	Very True (1)	Somewhat True (2)	A Little True (3)	Not At All True (4)
1. I put a lot of time and effort into my work.	○	○	○	○
2. The work made me want to find out more about the topic.	○	○	○	○
3. The directions were clear to me.	○	○	○	○
4. I felt involved in my work.	○	○	○	○
5. I liked what we did in science activity today.	○	○	○	○
6. I can use what I learned today later on.	○	○	○	○

7. The purpose of today's work was clear to me.	○	○	○	○
8. I was daydreaming about other things during the science activity.	○	○	○	○
9. I would like to do another activity like this sometime.	○	○	○	○
10. The work really made sense to me.	○	○	○	○

These sentences describe different reasons for doing schoolwork. Different kids have different reasons. We want to know how true each of the reasons was for why you did your science work.

Select the answer that best describes your reasons.

	A Lot Like Me (1)	Somewhat Like Me (2)	A Little Like Me (3)	Not At All Like Me (4)
1. I wanted to learn as much as possible.	○	○	○	○
2. I wanted to work with my friends.	○	○	○	○
3. It was important to me that the teacher thought I did a good job.	○	○	○	○

4. I wanted to do as little as possible.	○	○	○	○
5. I wanted to find out something new.	○	○	○	○
6. I wanted to talk with others about the work.	○	○	○	○
7. It was important to me to do better than other students.	○	○	○	○
8. I just wanted to do what I was supposed to and get it done.	○	○	○	○
9. It was important to me that I really understood the work.	○	○	○	○
10. I wanted to help others with their work.	○	○	○	○
11. I wanted others to think I was smart.	○	○	○	○
12. I wanted to do things as early as possible so I wouldn’t have to work very hard.	○	○	○	○

There are many different ways students do their work. We want to know how much each of these things are like what you did in science.

Select the answer that best describes your actions.

	A Lot Like Me (1)	A Little Like Me (2)	Not At All Like Me (3)
1. I followed the directions.	○	○	○
2. I tried to figure out how today's work fit with that I had learned before in science.	○	○	○
3. I guessed a lot so I could finish quickly.	○	○	○
4. I asked myself some questions as I went along to make sure the work made sense to me.	○	○	○
5. I wrote some things down.	○	○	○
6. I did my work without thinking too hard.	○	○	○
7. I explained or wrote down some things in my own words.	○	○	○
8. I checked to see what other kids were doing and did it too.	○	○	○
9. I paid attention to things I thought I was supposed to remember.	○	○	○
10. I skipped the hard parts.	○	○	○
11. I checked my science book or used other materials like charts when I wasn't sure about something.	○	○	○
12. I just did my work and hoped it was right.	○	○	○
13. I tried to figure out the hard parts on my own.	○	○	○
14. I copied down someone else's answers.	○	○	○

15. I went back over the things I didn't understand.	○	○	○

G.8 Open-ended Questions

1. What do you remember about the learning activity you just did?

2. What was hard to understand?

3. What was easy to understand?

4. What did you like?

5. What didn't you like?

6. Did you wish you had more time with Universe Sandbox? Why or why not?

G.9 Demographics

Gender?

- ○ Male
- ○ Female
- ○ Prefer not to identify

Age? ____

What is the last class or grade in which you remember learning about space?

____

What informal learning activities do you enjoy?

- ☐ Museums
- ☐ Zoos
- ☐ Aquariums
- ☐ Planetariums
- ☐ Interactive Science Centers
- ☐ Other(s) ______________________________________

Do you watch any of these television channels or shows?

- ☐ Cosmos
- ☐ NASA TV
- ☐ Discovery channel
- ☐ Science channel

☐ Bill Nye

☐ Other(s) ____

Do you follow other science news?

☐ On Facebook

☐ On Twitter

☐ Online news source

☐ Space magazines/websites

☐ Other(s)____

Have you ever attended a summer camp or a STEM (Science, Technology, Engineering, or Math) outreach day?

○ Yes

○ No

In the past, have you ever completed a research project for a science class on space?

○ Yes

○ No

Have you used Universe Sandbox before today?

○ Yes

○ No

○ Not sure

APPENDIX H. FOLLOW-UP INTERVIEW MATERIALS

H.1 Interview Questions

1. Can you describe what your experience with the solar system simulation was like?

    1A. What do you remember about it?

    1B. Can you describe a detail or two about the simulation?

2. Can you describe a detail or two about one of the planets?

3. When you answered the last few questions, how were you imagining or thinking about the solar system or the planets?

4. Did using the simulation help you think about the distance between planets (or the size of space in general) differently?

5. Looking back at the experience, was there something you liked?

6. Looking back at the experience, was there something you disliked?

7. Looking back at the experience, would you try something like this again?

8. Would you be excited to use something like this in school?

9. Is there anything else you'd like to share?

APPENDIX I. ADAPTED SURVEY MATERIALS

I.1 SAQ Original and Adapted

Two items were adapted in the SAQ Part 1 scale: items 5 and 8.

Item No.	Original	Adapted
5	I liked what we did in science today.	I liked what we did in the science activity today.
8	I was daydreaming about other things during science.	I was daydreaming about other things during the science activity.

I.2 MSLQ Original and Adapted

Twenty-six items were adapted for the MSLQ (all except for the Test Anxiety Subscale items: 3, 8, 14, 19, and 28).

Item No.	Original	Adapted
1	In a class like this, I prefer course material that really challenges me so I can learn new things.	In a learning activity like this, I prefer course material that really challenges me so I can learn new things.
2	If I study in appropriate ways, then I will be able to learn the material in this course.	If I study in appropriate ways, then I will be able to learn the material in this learning activity.
4	I think I will be able to use what I learn in this course in other courses.	I think I will be able to use what I learn in this learning activity in other courses.
5	I believe I will receive an excellent grade in this class.	I believe I will receive an excellent score in this learning activity.
6	I'm certain I can understand the most difficult material presented in the readings for this course.	I'm certain I can understand the most difficult material presented in the content for this learning activity.
7	Getting a good grade in this class is the most satisfying things for me right now.	Getting a good score in this learning activity is the most satisfying things for me right now.

9	It is my own fault if I don't learn the material in this course.	It is my own fault if I don't learn the material in this learning activity.
10	It is important for me to learn the course material in this class.	It is important for me to learn the material in this learning activity.
11	The most important thing for me right now is improving my overall grade point average so my main concern in this class is getting a good grade.	The most important thing for me right now is improving my overall score so my main concern in this learning activity is getting a good score.
12	I'm confident I can learn the basic concepts taught in this course.	I'm confident I can learn the basic concepts taught in this learning activity.
13	If I can, I want to get better grades in this class than most of the other students.	If I can, I want to get better score in this learning activity than most of the other students.
15	I'm confident I can understand the most complex material presented by the instructor in this course.	I'm confident I can understand the most complex material presented by the simulation in this learning activity.
16	In a class like this, I prefer course material that arouses my curiosity, even if it is difficult to learn.	In a learning activity like this, I prefer course material that arouses my curiosity, even if it is difficult to learn.
17	I am very interested in the content area of this course.	I am very interested in the content area of this learning activity.
18	If I try hard enough, then I will understand the course material.	If I try hard enough, then I will understand the learning activity material.
20	I'm confident I can do an excellent job on the assignments and tests in this course.	I'm confident I can do an excellent job on the assignments and tests in this learning activity.
21	I expect to do well in this class.	I expect to do well in this learning activity.
22	The most satisfying thing for me in this course is trying to understand the content as thoroughly as possible.	The most satisfying thing for me in this learning activity is trying to understand the content as thoroughly as possible.
23	I think the course material in this class is useful for me to learn.	I think the material in this learning activity is useful for me to learn.
24	When I have the opportunity in this class, I choose course assignments that I can learn from even if they don't guarantee a good grade.	When I have the opportunity in this learning activity, I choose activities that I can learn from even if they don't guarantee a good score.

25	If I don't understand the course material, it is because I didn't try hard enough.	If I don't understand the learning activity material, it is because I didn't try hard enough.
26	I like the subject matter of this course.	I like the subject matter of this learning activity.
27	Understanding the subject matter of this course is very important to me.	Understanding the subject matter of this learning activity is very important to me.
29	I'm certain I can master the skills being taught in this class.	I'm certain I can master the skills being taught in this learning activity.
30	I want to do well in this class because it's important to show my ability to my family, friends, employer, or others.	I want to do well in this learning activity because it's important to show my ability.
31	Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.	Considering the difficulty of this activity, the program, and my skills, I think did well in this learning activity.

APPENDIX I. PROCESS CODE DEFINITIONS

J.1 Interacting with Planet Information and Details

PC Codes	
Code Label	Code Description
pcOverviewTab	Exploring the overview tab
pcMotionTab	Exploring the motion tab
pcClimateTab	Exploring the climate tab
pcCompositionTab	Exploring the composition tab
pcPanelLarge	Leaving the large/full panel open during comparisons, or while looking at a range of planets
pcPanelSmall	Leaving the smaller info panel open during comparisons, or while looking at range of planets
infoPanelClose	Hiding the info panel specifically click off the side (pc)

VR Codes	
Code Label	Code Description
vrPointPlanet	Pointing controller at the planet to show the planet label/name

BOTH PC and VR Codes	
Code Label	Code Description
infoPanelOpen	Opening the main (little) info panel - either with hover (VR) or with click or focus change (in pc)
viewingMoons	Viewing the moons through zoom or teleport or from a distance around the planet (where the learner zoomed out and paused to watch the moons move)
viewingRings	Viewing the visible rings on Saturn (or re-teleporting to the planets with rings to hear them play in the audio conditions)
viewingSun	Viewing the Sun zoomed in or after teleport
zoomedViewing	Viewing a planet while zoomed in or after teleport

J.2 Using Movement or View Controls

PC Codes

Code Label	Code Description
pcZoom	Zooming in or out on a planet or the solar system
pcRotate	Rotating the view of the solar system
pcPlanetFocus	Setting a new planet focus (double clicking)
pcMoonFocus	Setting a new moon (double clicking)
pcSunFocus	Setting a new sun focus (double clicking)

VR Codes	
Code Label	Code Description
vrFly	Flying to change the view or to move inside of the system
vrGrip	Shifting the view of the solar system
vrTurn	Turning around physically/looking around in virtual environment
vrMoonTeleport	Teleporting to a moon
vrPlanetTeleport	Teleporting to a planet
vrSunTeleport	Teleporting to the sun
vrReset	Resetting the view to the starting location

BOTH PC and VR Codes	
Code Label	Code Description
pause	Pausing movement
restartProg	Closing and re-opening the program
changeView	Changing the view settings (e.g., removing the orbit lines)
changeTimeScale	Changing the time scale into something that's NOT paused

J.3 Verbalizing Comments and Observations

BOTH PC and VR Codes	
Code Label	Code Description
speakValue	Saying or reading numerical values
speakDetails	Saying or reading non-numerical values
speakPosOpinion	Saying any positive statement
speakNegOpinion	Saying any negative statement
speakUnexpected	Saying if something didn't meet their expectations (e.g., they thought one thing and were surprised)

speakCompare	Saying comparisons (e.g., facts about planets or whatever else)
speakExpected	Saying cases where they found a detail that confirmed something that they thought might be true
speakPattern	Describing any patterns (e.g., outside planets are colder, etc.) or things they noticed about the planets
speakSoundSearch	Talking about searching for a sound

J.4 Playing or Replaying Sounds

PC Codes	
Code Label	Code Description
pcPlaySound	Playing sounds on purpose (exploring while leaving the large panel open)
pcReplaySound	Re-selecting a planet to hear the rings sound

VR Codes	
Code Label	Code Description
vrPlaySound	Playing sounds on purpose (hitting the controller button when the panel is open)
vrReplaySound	Re-teleporting to a planet to hear the rings sound

J.5 Completing Non-Task Play

BOTH PC and VR Codes	
Code Label	Code Description
nonTaskPlay	Exploring and playing based on three different circumstances: 1. If they verbalize they're specifically looking at something outside of the exploration task they're completing 2. If it's at the end in the "free explore" time 3. If they do something like delete planets or change properties (PC only)

APPENDIX K. PROCESS CODING COUNTS AND STATISTICS

K.1 Code Counts

K.1.1 Group A Codes: PC – no-audio

	A201	**A202**	**A203**	**A204**
changesTimeScale	3	3	1	1
changeView	1	0	0	0
infoPanelClose	7	19	7	6
infoPanelOpen	73	69	69	116
nonTaskPlay	8	1	1	1
pause	0	3	1	1
pcClimateTab	0	1	0	9
pcCompositionTab	1	2	0	3
pcMoonFocus	1	0	4	1
pcMotionTab	5	1	0	7
pcOverviewTab	4	15	0	17
pcPanelLarge	2	8	0	3
pcPanelSmall	5	6	4	4
pcPlanetFocus	20	20	16	21
pcRotate	27	42	38	10
pcSunFocus	4	2	1	3
pcZoom	62	46	78	91

restartProg	3	0	0	0
speakCompare	5	29	3	0
speakDetails	11	16	11	1
speakExpected	2	15	2	0
speakNegOpinion	0	1	0	0
speakPattern	1	4	5	0
speakPosOpinion	4	1	0	0
speakSoundSearch	0	0	0	0
speakUnexpected	8	8	0	1
speakValue	0	21	0	0
viewingMoons	2	5	11	6
viewingRings	0	2	2	2
viewingSun	3	0	2	4
zoomedViewing	14	3	13	9
Totals	276	343	269	317

K.1.2 Group B Codes: PC – audio

	B301	B302	B303	B304
changesTimeScale	0	0	0	0
changeView	0	0	0	0
infoPanelClose	9	8	8	10
infoPanelOpen	44	108	91	50
nonTaskPlay	1	0	0	0
pause	0	0	0	0
pcClimateTab	1	7	1	0
pcCompositionTab	3	0	4	0
pcMoonFocus	0	0	0	0
pcMotionTab	2	0	9	0
pcOverviewTab	8	9	10	5
pcPanelLarge	3	2	3	2
pcPanelSmall	7	4	6	5
pcPlanetFocus	22	21	13	43
pcPlaySound	3	4	5	2
pcReplaySound	0	0	9	1
pcRotate	5	38	3	50
pcSunFocus	3	2	3	6
pcZoom	60	80	36	53
restartProg	0	0	0	0

speakCompare	0	16	0	7
speakDetails	0	21	0	14
speakExpected	0	4	0	9
speakNegOpinion	0	0	0	0
speakPattern	0	3	0	1
speakPosOpinion	0	0	0	0
speakSoundSearch	0	10	2	1
speakUnexpected	1	1	0	0
speakValue	0	0	1	20
viewingMoons	3	10	1	8
viewingRings	2	4	1	2
viewingSun	2	0	1	1
zoomedViewing	17	7	4	6
Totals	196	359	211	296

K.1.3 Group C Codes: VR – no-audio

	C401	**C402**	**C403**	**C404**	**C405**
changesTimeScale	1	0	0	1	0
changeView	2	0	0	1	1
infoPanelClose	0	0	0	0	0
infoPanelOpen	44	65	51	57	27
nonTaskPlay	1	1	1	1	1
pause	3	2	0	2	0
restartProg	1	0	0	0	0
speakCompare	4	2	16	2	0
speakDetails	19	17	5	8	6
speakExpected	2	0	3	0	0
speakNegOpinion	0	0	4	0	0
speakPattern	2	1	13	0	0
speakPosOpinion	0	1	1	1	0
speakSoundSearch	0	0	0	0	0
speakUnexpected	2	0	2	0	0
speakValue	5	0	9	2	0
viewingMoons	7	0	2	6	0
viewingRings	4	0	2	0	0
viewingSun	0	0	0	1	0
vrFly	0	4	3	5	0

vrGrip	0	36	4	0	0
vrMoonTeleport	29	38	13	16	9
vrPlanetTeleport	24	42	19	41	20
vrPointPlanet	141	138	48	93	61
vrReset	13	1	32	0	12
vrSunTeleport	8	9	0	4	1
vrTurn	120	163	69	81	61
zoomedViewing	0	0	0	2	0
Totals	432	520	297	324	199

K.1.4 Group D Codes: VR – audio

	D501	D205	D503	D504
changesTimeScale	3	6	3	1
changeView	0	0	0	0
infoPanelClose	0	0	0	0
infoPanelOpen	62	55	58	45
nonTaskPlay	1	1	1	1
pause	11	6	6	8
restartProg	0	0	0	0
speakCompare	14	16	8	8
speakDetails	17	12	13	3
speakExpected	9	2	0	1
speakNegOpinion	0	0	1	0
speakPattern	4	4	1	2
speakPosOpinion	3	0	3	0
speakSoundSearch	1	5	1	0
speakUnexpected	5	0	4	1
speakValue	57	9	8	1
viewingMoons	0	2	6	0
viewingRings	1	0	4	0
viewingSun	0	0	0	0
vrFly	10	1	0	0

vrGrip	0	1	9	3
vrMoonTeleport	16	24	20	16
vrPlanetTeleport	22	49	44	27
vrPlaySound	2	50	7	0
vrPointPlanet	60	109	84	55
vrReplaySound	2	25	12	5
vrReset	21	0	3	8
vrSunTeleport	1	2	1	1
vrTurn	124	151	130	116
zoomedViewing	4	0	0	0
Totals	450	530	427	302

K.2 Chi-square Observed

K.2.1 Top-level Code Categories

Observed values for top-level code categories.

	Planet Info Details	Movement Controls	Verbalizations	Rotating & Turning	Totals
Group A	541	504	149	117	1311
Group B	488	438	111	96	1133
Group C	749	877	127	494	2247
Group D	545	800	213	521	2079
Total	2323	2619	600	1228	6770

Degrees of Freedom (df): 9, $p < .001$

Test statistic: 318.23

Critical value = 16.92

Standard residuals (z-scores); note: bold scores are for where differences occurred.

	Planet Info Details	Movement Controls	Verbalizations	Rotating & Turning
Group A	**4.30**	-0.14	**3.04**	**-7.83**
Group B	**5.03**	0.01	1.06	**-7.64**
Group C	-0.79	0.26	**-5.11**	**4.28**
Group D	**-6.30**	-0.15	**2.12**	**7.41**

K.2.2 Viewing Details Only

Observed values for codes which related to the viewing details.

	Info Panel Open	Viewing Moons	Viewing Rings	View Planet	Totals
Group A	327	24	6	39	396
Group B	293	22	9	34	358
Group C	244	15	6	146	411
Group D	220	8	5	142	375
Total	1084	69	26	361	1540

df: 9, $p < .001$

Test statistic: 233.66

Critical value = 16.92

Standard residuals; note: bold scores are for where differences occurred.

	Info Panel Open	Viewing Moons	Viewing Rings	View Planet
Group A	-1.94	0.17	-0.92	**-4.47**
Group B	**-1.99**	0.23	0.39	**-7.22**
Group C	**-6.89**	-1.85	-1.02	1.80
Group D	**-6.72**	**-2.98**	-1.14	**2.54**

K.2.3 PC-Only Comparisons

Observed values for codes which related to the viewing details for the PC conditions.

	Climate Tab	Composition Tab	Motion Tab	Overview Tab	PC Panel Large	PC Panel Small	Planet Focus	Total
Group A	10	6	13	36	13	19	77	174
Group B	9	7	11	32	10	22	99	190
Total	19	13	24	68	23	41	176	364

df: 6, p = 0.78

Test statistic: 3.20

Critical value = 12.59

Standard residuals were not calculated as there were no differences between the PC - conditions for the viewing details.

K.2.4 VR-Only Comparisons

Observed values for codes which related to movement and label viewing for VR conditions.

	VR Point Planet	VR Fly	Grip	Moon Tele.	Planet Tele.	Reset	Sun Tele.	Turn	Pause	Total
Group C	481	12	40	105	4146	58	22	494	7	1365
Group D	308	11	13	76	142	32	5	521	31	1139
Total	789	23	53	181	288	90	27	1015	38	2504

Degrees of Freedom (df): 8, $p < .001$

Test statistic: 70.70

Critical value = 15.51

Standard residuals; note: bold scores are for where differences occurred.

	VR Point Planet	VR Fly	Grip	Moon Tele.	Planet Tele.	Reset	Sun Tele.	Turn	Pause
Group C	**2.45**	-0.15	**2.07**	0.64	-0.88	1.28	1.90	**-2.52**	**-3.01**
Group D	**-2.69**	0.17	**-2.26**	-0.70	0.96	**-2.08**	**-2.08**	**2.76**	**3.30**

K.2.5 Verbalizations

Observed values for codes which related to verbalizations.

	Speak Expected	Speak Compare	Speak Details	Speak Value	Total
Group A	19	37	39	21	116
Group B	13	23	35	21	92
Group C	5	24	55	16	100
Group D	12	46	45	75	178
Total	49	130	174	133	486

Degrees of Freedom (df): 9, $p < .001$

Test statistic: 51.50

Critical value = 16.91

Standard residuals; note: bold scores are for where differences occurred.

	Speak Expected	Speak Compare	Speak Details	Speak Value
Group A	**2.14**	1.07	-0.39	-1.91
Group B	1.22	-0.32	0.35	-0.83
Group C	-1.60	-0.53	**3.21**	**-2.17**
Group D	-1.40	-0.23	**-2.35**	**3.77**

APPENDIX L. POST-ACTIVITY FREE RESPONSE

Note that all responses are quotes.

L.1 Question 1

What do you remember about the learning activity you just did?

A201	*I deleted Earth* *I turned mars into light* *There are no baseballs* *Venus is the hottest even thogh mecury is the closest to the sun*
A202	*mostly it was the which planets had moons, and rings*
A203	*How to say most of the planets in English.* *That many planets have moon.*
A204	*The learning activity dealt with the solar system and different characteristics of each planet, such as mass, density, speed, location, and composition.*
B301	*mercury is the closest plant to the sun*
B302	*It was enjoyable, interesting, and simple*
B303	*I remember clicking on different planets looking at and hearing different statistics for each.*
B304	*I remember the detail of the planets and the interface of the program, which showed the mass of the planets as well as important information such as their temperature and size measured in simpler terms (such as one earth, moon, or sun).*
C401	*That it was a lot of information about planets*
C403	*Seeing all the planets and some of the weird orbits some of the moons had*
C403	*I remembered that only Jupiter and Saturn have rings and that Venus and Earth have similar gravitational strength.*
C404	*I took a trip around the universe and learned some things about the planets in our solar system.*
C405	*I remember using a virtual reality technology to explore the solar system and the planets*
D501	*I learned that venus is one of the hottest planets in the solar system*

D502	*Getting to see the planets up close and learning information about them. I was able to see Earth from space during day and night, which was interesting.*
D503	*Exploring parts of the solar system. Explored all the different planets and found out details that I before did not know.*
D504	*The planet's rotation were similar to their actual rotation. The hottest planet is Venus and the coldest is Uranus*

L.2 Question 2

What was hard to understand?

A201	*Why there are no baseballs*
A202	*it was all pretty easy to understand*
A203	*The mass of the planets.*
A204	*Some of the terms in the info boxes were unknown to me, so it was a bit difficult to understand those parts.*
B301	*no*
B302	*No, it was pretty easy*
B303	*I was a little confused about the sounds.*
B304	*Nothing that I can remember.*
C401	*The questions about the planets because I haven't studied on planets for awhile*
C403	*The gravitational strength of the planets*
C403	*To see how fast the planets rotated for some of the ones with close rotation times.*
C404	*Some of the questions were hard to answer because I haven't used this program before.*
C405	*Nothing*
D501	*the random sounds the planets made*
D502	*Not much. I think the information and sounds were clear enough to me.*
D503	*It was hard to understand some of the sounds sometimes.*
D504	*The overall assignment was easy to understand, but the sounds were a little hard to understand*

L.3 Question 3

What was easy to understand?

A201	*how to rename a whole planet*
A202	*mostly the moons and rings, which is why I remembered it so well*
A203	*The names and the order.*
A204	*The visual aspect was easy to play around with and notice things. The overview of each planet gave information that was easy to understand.*
B301	*yes*
B302	*Yes*
B303	*It was simple to navigate.*
B304	*The interface, as it showed a lot of information in a very compact and neat little box*
C401	*No*
C403	*The mass of the planets*
C403	*Where the planets are and the speed of revolution.*
C404	*Most of the questions were easy to understand once I figured out what the meant.*
C405	*How do travel around and look at the planets*
D501	*the movements on the controller*
D502	*The information displayed about a planet and the sounds that went along with it.*
D503	*The concept of what was supposed to happen and how to use it.*
D504	*Like I previously stated, The overall assignment was easy to understand, but the sounds were a little hard to understand*

L.4 Question 4

What did you like?

A201	*Deleting earth* *Renaming earth to planet* *Turning mars into light* *Blowing up the entire universe*
A202	*All that was included in the test*
A203	*I liked that I got to explore the planets that I didn't know.*
A204	*I liked being able to change the perspective of the solar system so that I could view it from different angles. Also, I liked being able to zoom in and out on the planets to do examinations.*
B301	*supernova*
B302	*I liked learning more about the solar system*
B303	*I really like how real it feels and the atmosphere.*
B304	*The detail of the planets and the customization that's available (if given the opportunity, as none of the planet's aspects were tweaked at all)*
C401	*The VR and the directions*
C403	*Exploring the solar system*
C403	*I loved the feeling of actually being there. It felt so realistic and it felt like I could actually touch the planets and stars.*
C404	*I liked being in virtual reality I general but I also liked how I could fly and teleport around to different things.*
C405	*I liked that I could see information about each planet*
D501	*all that was included in the test*
D502	*I liked the option of pausing time, as it made navigation easier. Being able to explore the moons alongside the planets was fun.*
D503	*I liked how it showed facts of the solar system in a interesting way and it showed the facts while it was interactive/hands on.*
D504	*The fact the the sun was really bright and the fact that you can try to catch*

L.5 Question 5

What didn't you like?

A201	*writing*
A202	*bummed out that it was just on the computer – imagining a VR headset, but not too much. Would have rather used a VR version*
A203	*nothing much*
A204	*I didn't like how you had to double click on a planet to center it because I would keep forgetting to double click.*
B301	*nothing*
B302	*That it seemed like it was over so quickly.*
B303	*I am not sure*
B304	*The zooming and navigation was a little finicky at times, but it didn't inhibit my ability to learn*
C401	*Nothing really*
C403	*It was a little difficult to grasp where things were in relation to each other*
C403	*It was a little hard to click on then planets because the moon were usually in the way of it.*
C404	*Nothing really but when you take off the headset you get a weird feeling around your eyes.*
C405	*I didn't like that the planets were moving*
D501	*i did not dislike any thing*
D502	*I don't think there is anything I don't like about the Universe Sandbox.*
D503	*Sounds were a little confusing.*
D504	*The Audio information*

L.6 Question 6

Did you wish you had more time with Universe Sandbox? Why or why not?

A201	*yes* *So I can destroy more stuff* *and find the baseballs*
A202	*no, thinks it was enough time. It was all very straight to the point. Summarized the planet in a way that I could understand it.*
A203	*Yes because you can see all the planets and especially the moons.*
A204	*I wish that I had more time so that I could figure out what some of the terms meant. I also would like to figure out the answers to the questions that I didn't know. In addition, I wanted to play around with the simulation for a little while longer.*
B301	*yes because it was fun*
B302	*Yes because I find it very fun to use and would like to learn more.*
B303	*I do not really wish for more time with Universe Sandbox because I feel that I have explored enough that I am satisfied.*
B304	*Yes, because I'd like to see what happens when I begin messing with the temperatures and size of the planets outside of their original values, as it'd be interesting to see how it affects the rest of the Solar System.*
C401	*Yes so I could keep exploring*
C403	*Yes. I liked exploring and want to explore some more.*
C403	*Yes, I had a really great time and I wanted to learn more about it. It was also very cool because you could really see the planets and it felt like you are really in space.*
C404	*I do wish I had more time with a different scene because its so fun to just be around in virtual reality cause you an mess up or do something bad and its not real.*
C405	*Yes because it was fun*
D501	*Yes, it is so fun the picture looks so real and I learned a lot from this*
D502	*Yes, because I enjoyed looking at a scale of the Solar System.*
D503	*Yes because it taught information in a hands on way to make it easier to understand.*
D504	*Yes. I wanted to see if after awhile I could understand the audio information. Also I wanted to try to catch all the fast rotating planets*

APPENDIX M. FOLLOW-UP INTERVIEW RESPONSES

Note that all responses are quotes unless otherwise stated.

M.1 Question 1

Can you describe what your experience with solar system simulation was like?

A202	First off it wasn't boring it was kinda fun, being able to click on the planets and find out what they were. The visual representation of the moons and the rings was cool. And, in general it was really cool.
A203	It was fun cause I've got that website -- the sandbox thing, to look at the moons and stuff and how it's moving.
A204	It was kinda cool playing around and being able to see all of the things about the planets and being able to learn the different things
B302	Um, there were different sounds that meant different things. The different things were like mass, or moons, rings, other kinds of stuff for different planets about the solar system.
B303	It was pretty cool, like you could see all the planets around it.
B304	Uhm, it was pretty cool, I think.
C401	It was pretty fun. It felt so real, like I was actually in the solar system
C402	It was like being in space, but like, I don't know -- it was, I felt like I was there.
C403	Well, I had a really good time doing it. It really did feel like I was kind of in space, and just looking at the planets. IT was just really fun and it was a great learning experience for me.
C404	We were in, was in a space thing and there were lots of planets and some moons, too.
C405	It was like, kind of like being there, because you couldn't see myself, but I could see all around, instead of just in one little spot. And, like, there were planets moving and stuff.
D502	Hmm… I think it, to me, was, it interested me. I wanted to explore it. So. I don't know if there's a word for it.
D503	I remember seeing all of the planets in a 3d looking area. Like being able to see, like, learning it, kind of like learning hands on, cause it was 3d and you could interact with it.
D504	So, it was really really awesome having to actually look at the other planets and stuff. I left feeling like it was a really awesome experience.

	The hand mechanics made sense, but when it came to the sound mechanics, I had trouble keeping up with a few of them. Other than that, I don't think I had any difficulties keeping up with it.

M.2 Question 2

What do you remember about it?

A202	(no response)
A203	So, I remember the closest planet to the sun is Mercury, Venus, Earth, and then Mars. Jupiter, Saturn, and then…like Neptune
A204	I remember the side bars which had info like the mass and composition
B302	I thought it was pretty cool. Just like, going around space, hearing sounds to know different things, and I liked not having to spend all of the time reading, b/c I'm a slow reader.
B303	There were different sounds for each planet, like, describing it's mass, density, and other features.
B304	I remember it was the solar system you could see all of the planets, orbiting around the sun and all. You could see information like the mass, average temp, etc.
C401	I'm going to say not a lot -- (about the questions). I remember looking at all of the planets, like the sun or mercury, or how you could stop time.
C402	I remember seeing all of the planets and the crazy orbits of the moons and stuff.
C403	I remember that it looked like you were surrounded by the milky way. There was all the planets and their moons, or some of their moons, for the most part.
C404	Well, I know you could use special controls to like move around and stuff, and it was really big and all of the planets were there but not all of the moons.
C405	I remember that you could teleport to the different planets and it gave you information about the mass and temperature, and how many moons. And you asked me questions about which ones had rings, and which ones were bigger than the others.
D502	I think the most memorable part was probably the sounds you heard upon clicking upon a planet.
D503	And like, you learned a lot of information about the planets.
D504	I think, I remember the most about the actual speed -- the speed was comparable to what it would be like, in actual terms. But because of their size -- it appeared really fast. B/c of their actual size (in VR) it was smaller in diagram. It was moving a lot faster, and it was pretty awesome. It was fun to try to catch them without pausing. The teleportation mechanic was cool cause you can view them and go directly to the planets

M.3 Question 3

Can you describe a detail or two about the simulation?

A202	The first detail is that I remember that saturn had both rings and moons. And the sun was made of hydrogen.
A203	Like, how much degrees or how big in masses and stuff.
A204	There were parts where you could zoom in and see the moons and their motion around the planets
B302	I remember being able to scroll out and see all of the planets, or scrolling in to focus on a singular one. Like, when I was further out I could hear all of the masses about the planets, but when I was in, I could hear just the one.
B303	Each planet had certain information about them, like the specific numbers for each category of information.
B304	I guess, I remember some of the questions, what two planets have similar masses. I remember it was Neptune and Venus having similar mass.
C401	Some of the details that stood out were that it could actually be used for learning. Thought it was really cool, and thought it would be awesome if you could use this in schools.
C402	I remember, I think, one of the moons of Jupiter, or Uranus had a loopy orbit.
C403	So, the simulation in general, you could click on the planets and could see the details or facts about them.
C404	It was kinda like a simulation of the galaxy and there were lots of planets and it made it seem really big
C405	Uhm, the planets were moving, and they were like, really awesome, they were sort of realistic, kind-of.
D502	Talking details, then I do remember when I looked at the sun, it was pretty bright. The sun was the first thing I clicked on.
D503	There was, so there was 2 modes -- fly and teleport, and I used teleport the whole time. And, uh, there were sounds to go with each fact about the planets.
D504	I do remember that Jupiter, had the most moons. Saturn had the most obvious rings. From looking at it. Looking at it, neptune was the coldest planet. The sun -- it was really really bright, but when you looked away it dimmed down.

M.4 Question 4

Can you describe a detail or two about one of the planets?

A202	Remember that, I believe it was mercury and mars gravitational strengths were mostly the same.
A203	I think the earth was the hottest. Oh, um, and the planets are closest to the sun they're kind of hot and if they're not it's so cold.
A204	I remember Saturn was the only one with rings. They looked like little tiny black dots around it
B302	Jupiter was the densest gas giant, Saturn had rings, Uranus was like -- they all had moons except for Mercury and Venus.
B303	There was like, pretty sure like, I remember something about the rotation speeds, but I don't remember which specific planets. It sounded pretty interesting.
B304	Just like -- I remember having Jupiter having quite a few moons and all, spinning around, it was cool to see.
C401	[One that stood out to him was the sun, how bright it was.] "If you got up close to it, it got brighter. being able to go on the sun in the VR would be cool." [additionally mentioned, un-specifically, how the planets looked. Would like it to be more detailed.]
C402	I remember that Jupiter and it was a gas giant, and that Saturn was the only one that had rings. And mercury was closest to the sun.
C403	Yeah. The main thing I was remembering was just the rings on Saturn and Jupiter. But I don't really remember any more details about them.
C404	When you clicked on, for Mars, or Venus, there was something about the center of gravity, or some gravitational pull or force. And it was complicated --one of the questions -- and you could see the stuff on the side
C405	Well, I know that, like, I knew that Saturn had rings, and Saturn and Jupiter were the biggest. And that, uhm, Saturn and Jupiter are moderately, have moons. I learned that Neptune had a moon.
D502	The trails that the planets left behind. The noise, that, the sounds that -- Saturn's rings, or any of the rings. Saturn's stood out b/c it was both visual and sound. Sometimes the planets, but mainly earth, would get light or dark, depending on the time.
D503	So, uh…Jupiter was the largest planet with the biggest mass. And, it has a lot of moons
D504	Mercury at first glance looked comparable to the moon -- it looked all dry and no life and color at all. While, liked planets like Venus and earth had more texture (and nothing on mars) -- Venus appeared different than I thought it would be (it was a yellow-ish tan) as opposed to the amount of red -- light if it had lava. All the blinking lights around it.

M.5 Question 5

When you answered the last few questions -- How were you imagining or thinking about the solar system or the planets?

A202	It was mostly the planets that I paid more attention to, or looked more interesting. Or that I've looked at over the years, like mars. In my robotics, I believe, I think we researched life on mars. We were researching that there might be bacteria in the water. So I paid a lot of attention to that (the sun) b/c it's a big ball of fire.
A203	Entire solar system and like looking at the planets, they were like so big.
A204	Kinda remember how the planets looked
B302	I kind of, saw it, as the simulation showed it. (thought about all of them in general, and then pictured one in particular while he was describing a detail)
B303	I imagined it like how the simulation displayed it, I guess, in a ring format.
B304	Um, I kinda imagined what it was like in the simulation.
C401	Thinking of them like physically -- that you could actually physically touch them.
C402	I was thinking about it like how I saw it in that program
C403	Thinking about them like probably physically in there. You could imagine you were there.
C404	I think, I guess it looked a lot like the simulation
C405	I pictured them, from like, in the simulation.
D502	As in like, how did I picture it in my mind? Probably individually with the planet in the center.
D503	I was thinking about how this planet, what it is, and kinda learned about them.
D504	Thought about the top-down view overall. Looking at them, I kinda like, when we were at the individual -- I would start visualizing the planets and the diagrams that were shown.

M.6 Question 6

Did using the simulation help you think about the distance between planets (or the size of space in general) differently?

A202	Yeah, cause some planets like mercury and Venus were closer together, but the gas giants, were far apart, like how Jupiter and Saturn were far apart, like the orbital lines helped tell how close or far apart they were.
A203	Oh yeah, yeah definitely. -- I didn't know that like, Mars and Jupiter or Jupiter and Saturn were so far away from each other. It was spinning so fast (their spins)
A204	Um, when I first saw it and saw how far away the outside planets, Uranus and Neptune were, I was shocked cause the ones closer to the sun were close together
B302	It did a bit, but… it helped me with being able to see the sizes of the planets and the distance some, but not as much, with how far they were. (there could still be something else to better understand the distance)
B303	Not really. It helped get it into perspective, but I kinda already knew.
B304	Uhm, I think so. Originally it's like yea it's this far away (light years away) for planets -- all the planets being far apart. I struggled to navigate to all of them, since they were so far apart.
C401	Oh yeah it did. Thought some of the planets were pretty small -- but in the VR you can see how big they actually are.
C402	Yeah. I could see like, how big they were in relation to each other, or how *really* far apart they were.
C403	Yeah, you could see the differences like how the rocky planets were, way closer together compared to the gas giants. (to the sun)
C404	Yeah, I think, having to teleport to each planet made me realize it was vigger, you couldn't just fly to them. Even the center planets weren't as close.
C405	I realized that they were a lot further away from each other. I guess. They were really spaced apart.
D502	Distance between the planets? Yes. I could definitely see the distance.
D503	So, the simulation allowed, to show like, it made me able to see how these planets are actually a lot further or a lot closer together than I used to think they are. Like how the farther planets are a lot more spaced out.
D504	It actually did make me think more about the distance, because in my head, in the diagrams and images, the planets are really close together (back to back to back) but in VR I could see the rocky giants and then the space -- like Uranus all the way in the very back.

M.7 Question 7

Looking back at that experience, was there something you liked?

A202	I liked how you could click on the planets and it would bring you closer to it. And how you could just see the planet up close, even though you're sitting down in a room on one planet, where you're nowhere near the other planets that you're looking at.
A203	Yeah, so like you can double click and see the planets closer and you can see the moons closer, too. I like that.
A204	Um, kinda liked playing around with it (at the end)
B302	I liked, really liked science, and being able to interact with it a lot. Cause, I'm more of a hands-on learner and have fun with that.
B303	Like you could look around and see the planets spinning, and the moons as well
B304	Um, I really liked that you can see the run down of the planets if you clicked on them.
C401	That you could do anything with the planets (like shrink the sun) [the view]
C402	I liked exploring space and looking at the planets and stuff. I liked being in the thing.
C403	I really liked that I learned a lot from it. I really liked that it was just, the whole experience in general.
C404	Uhm, the way you could move around (look around)
C405	I liked that I could, uhm, that I could go to each planet and see all of the information about it. And that I could see what's big, and what had rings and moons, and stuff.
D502	It was nice to go to each planet and look at it up close.
D503	I liked how you could look around the solar system as if you were there. And, being able to see each planet
D504	I did list this on the paper: I absolutely loved the planet mechanic (how fast they moved -- it gave it a lot of excitement). This was really entertaining to catch the planets and moons in mid-flight. The sound mechanic was enjoyable but was confusing at times. I remember than when you clicked on a planet, you could hear the mass and density. the lower the lower the mass or density.

M.8 Question 8

Looking back at that experience, was there something you disliked?

A202	Not really actually.
A203	Not really.
A204	Uhm, I guess like having to do the questions over and over.
B302	Not really, no.
B303	Uh, I can't really say anything.
B304	Uhm, not really sure.
C401	No, didn't dislike it. It was actually pretty good
C402	Um… not really.
C403	Not just me, but when I tried to click on the far away planets, I usually clicked on the moons, not the planets. There's probably an easier way to do it. Everything else was really awesome.
C404	I guess, I didn't like, when you had to fly. It wasn't going really fast. I liked the rest of it.
C405	No, I don't think. I liked it all.
D502	I don't think so.
D503	Sometimes the sounds were a little bit confusing.
D504	The sounds -- it was confusing.

M.9 Question 9

Looking back at that experience, would you try something like this again?

A202	Yeah, usually I would, b/c it's a pretty big step from my learning science class. Usually my teacher would have us do something similar to this, called a gizmo, it's not as visual as yours was, it was more of, you just click a few buttons and it tells you something. It looks more 2D and yours was more 3D.
A203	Yeah.
A204	Yeah
B302	Yeah -- definitely.
B303	Yeah
B304	Sure, probably.
C401	Yes, I would
C402	Yeah!
C403	Yeah, definitely. Yeah. It was really cool.
C404	Yeah, definitely.
C405	Probably.
D502	Yes, I would.
D503	Yeah
D504	Most definitely, I would. It was really fun.

M.10 Question 10

Would you be exited to use something like this in school?

A202	Oh yeah, it would make science or other subjects a lot more interesting.
A203	Mhm, it will help my friends a lot, probably.
A204	Yeah - it's better than just doing a web class, it was more engaging and interactive.
B302	Yeah - I think it'd be fun.
B303	Yeah, it would be pretty cool. Yeah.
B304	Mhm, definitely. I guess it helps cause you can actually interact with all of the planets. It helps you remember a bit more, seeing the detailed view, in a place you can see. I actually haven't gone over the planets in years -- here's this planet, with this much mass. You could actually look around and view all the planets.
C401	Yeah, I would be really excited.
C402	Yeah, I think that would be awesome to use it in school.
C403	Definitely. That would be a great way to learn, in class periods.
C404	Yeah, I'd like it a lot. Like, for not just science, any subject.
C405	I think it would be really fun. I liked that you could explore around and visualize it better than looking at a picture in the textbook.
D502	Yeah, I'd be looking forward to it.
D503	Yeah, it would be easier to use, learning is hands on.
D504	Yes- it would make the classroom so much more fun.

M.11 Question 11

Is there anything else you would like to share?

A202	Not really.
A203	It was the sun, Mercury, Venus, Earth, Mars, Jupiter, Saturn, and then the last one was Neptune, but… I knew it yesterday, uh.
A204	uhm, nope.
B302	Nope. Not that I can think of.
B303	Nothing in particular.
B304	Some of them I could get (the sounds); I didn't really use or utilize the sounds that much. I think I remember mercury having a fast-paced high-pitched sound.
C401	Nope
C402	Nope
C403	I don't think so.
C404	I don't remember anything else. I felt there were lines going through the planets around them -- lots of lines, connecting to the planets and the moons (the orbits).
C405	Um, I guess that I liked it.
D502	Hmm. No I think that's all.
D503	Hm, no.
D504	I must say that I can see the amount of hard work you put into it, the diagrams and the planets, I was really impressed. Thank you.

REFERENCES

Abrahamson, D., & Sánchez-García, R. (2016). Learning Is Moving in New Ways: The Ecological Dynamics of Mathematics Education. *Journal of the Learning Sciences*, *25*(2), 203–239. https://doi.org/10.1080/10508406.2016.1143370

Achieve, I. (2013). *DCI Arrangements of the Next Generation Science Standards* (Issue April).

Ackermann, E. (2001). Piaget's Constructivism, Papert's Constructionism: What's the difference? *Constructivism: Uses and Perspectives in Education*, *5*(3), 1–11. https://doi.org/10.1.1.132.4253

Adams, W. K., Perkins, K. K., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2005). The Design and Validation of the Colarado Learning Attitudes about Science Survey. *2004 Physics Education Research Conference*, 45–48. https://doi.org/10.1063/1.2084697

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics - Physics Education Research*, *2*(1), 1–14. https://doi.org/10.1103/PhysRevSTPER.2.010101

Adams, W. K., Reid, S., Lemaster, R., McKagan, S. B., Perkins, K. K., Dubson, M., & Wieman, C. E. (2008). A Study of Educational Simulations Part 1 – Engagement and Learning. *Journal of Interactive Learning Research*, *19*(3), 397–419. http://avoserv.library.fordham.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2009-01931-002&site=eds-live%5Cncwieman@jila.colorado.edu%5Cnmichael.dubson@colorado.edu%5Cnkatherine.perkins@colorado.edu%5Cnsarah.mckagan@colorado

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (2016). From First Grade Forward : Early Foundations of High School Dropout. *American Sociological Association*, *70*(2), 87–107.

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, *26*(1063519), 1–13. https://doi.org/10.25304/rlt.v26.2140

Allen, S. (2004). Designs for learning: {Studying} science museum exhibits that do more than entertain. *Science Education*, *88*(S1), S17--S33. https://doi.org/10.1002/sce.20016

Alves, V., & Roque, L. (2010). Guidelines for sound design in computer games. *Game Sound Technology and Player Interaction: Concepts and Developments*, 362–383.

https://doi.org/10.4018/978-1-61692-828-5.ch017

Ament, V. T. (2014). *The Foley grail: The art of performing sound for film, games, and animation*. CRC Press.

Andreas Schleicher. (2019). *PISA 2018 insights and interpretations*. https://www.oecd.org/pisa/PISA 2018 Insights and Interpretations FINAL PDF.pdf

Antle, A. N., Droumeva, M., & Corness, G. (2008). Playing with the sound maker: do embodied metaphors help children learn? *Proceedings of the 7th International Conference on Interaction Design and Children - IDC '08*, 178–185. https://doi.org/10.1145/1463689.1463754

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*(5), 369–386. https://doi.org/10.1002/pits.20303 STUDENT

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Archer, L., Dawson, E., Seakins, A., DeWitt, J., Godec, S., & Whitby, C. (2016). "I'm Being a Man Here": Urban Boys' Performances of Masculinity and Engagement With Science During a Science Museum Visit. *Journal of the Learning Sciences*, *25*(3), 438–485. https://doi.org/10.1080/10508406.2016.1187147

Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers and Education*, *53*(4), 1339–1354. https://doi.org/10.1016/j.compedu.2009.06.017

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), 136–140. https://doi.org/10.1016/j.cub.2009.12.014

Bakker, S., Antle, A. N., & van den Hoven, E. (2009). Identifying embodied metaphors in children's sound-action mappings. *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 140. https://doi.org/10.1145/1551788.1551812

Ballesteros, J. C., & Luque Serrano, B. (2008). *Using sounds and sonifications for astronomy outreach*. http://oa.upm.es/4724/

Ballora, M. (2014). Sonification Strategies for the Film Rhythms of the Universe. *International Conference on Auditory Display*.

https://smartech.gatech.edu/handle/1853/52075

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123. https://doi.org/66.39.39.113

Barab, S. A., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology*, *16*(1), 1–3. https://doi.org/10.1007/s10956-007-9043-9

Barab, S. A., Hay, K. E., Barnett, M., & Keating, T. (2000). Virtual solar system project: Building understanding through model building. *Journal of Research in Science Teaching*, *37*(7), 719–756.

Barg-Walkow, L. H., Mitzner, T. L., & Rogers, W. A. (2014). *The Technology Experience Profile (TEP): Assessment and Scoring Guide*.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with Understanding : Lessons from Research on Problem- and Project-Based Learning. *The Journal of the Learning Sciences*, *7*(3), 271–311.

Barros, R. M., Silver, E. J., & Stein, R. E. K. (2009). School Recess and Group Classroom Behavior. *Pediatrics*, *123*(2), 431–436. https://doi.org/10.1542/peds.2007-2825

Batterman, J. M., Schuett, J. H., & Walker, B. N. (2013). Auditory displays for accessible fantasy sports. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '13*, *1*(404), 1–2. https://doi.org/10.1145/2513383.2513397

Batterman, J. M., & Walker, B. N. (2012). Displaying error & uncertainty in auditory graphs. *Proceedings of the 14th International {ACM} {SIGACCESS} Conference on {Computers} and Accessibility*, 285–286. https://doi.org/10.1145/2384916.2384995

Bédard, D., Lison, C., Dalle, D., Côté, D., & Boutin, N. (2012). Problem-based and project-based learning in engineering and medicine: Determinants of students' engagement and persistence. *Interdisciplinary Journal of Problem-Based Learning*, *6*(2), 7–30. https://doi.org/10.7771/1541-5015.1355

Beer, J. M. (2013). *Understanding Older Adults' Perceptions of Usefulness of an Assistive Home Robot.* (Issue December) [Georgia Institute of Technology]. https://doi.org/10.1073/pnas.0703993104

Betts, J. E., Appleton, J. J., Reschly, A. L., Christenson, S. L., & Huebner, E. S. (2010). A study of the factorial invariance of the Student Engagement Instrument (SEI): Results from middle and high school students. *School Psychology Quarterly*, *25*(2), 84–93. https://doi.org/10.1037/a0020259

Billinghurst, M., & Dünser, A. (2012). Augmented reality in the classroom. *Computer*, *45*(7), 56–63.

Billinghurst, M., & Kato, H. (2002). Collaborative augmented reality. *Communications of the ACM*, *45*(7), 64–70.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating Project-Based Learning: Sustaining the Doing, Supporting the Learning. *Educational Psychologist*, *26*(3–4), 369–398. https://doi.org/10.1080/00461520.1991.9653139

Bonebright, T. L., Nees, M. A., Connerley, T. T., & McCain, G. R. (2001). *Testing the effectiveness of sonified graphs for education: A programmatic research project*.

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., & Wilderman, C. (2009). Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. *Online Submission*, *July*.

Bormann, K. (2005). Presence and the Utility of Audio Spatialization. *Presence: Teleoperators and Virtual Environments*, *14*(3), 278–297. https://doi.org/10.1162/105474605323384645

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *In How People Learn: Brain, Mind, Experience, and School (Expanded edition)*. National Academies Press. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C11&q=how+people+learn+brain+mind+experience+and+school&btnG=

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer : A Simple Proposal with Multiple Implications. *Review of Research in Education*, *24*(1999), 61–100.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Bregman, A. S. (1993). Auditory scene analysis: {Hearing} in complex environments. *Thinking in Sound: The Cognitive Psychology of Human Audition*, 10–36.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*, *April*, 222–227. https://doi.org/10.1145/169059.169179

Brooke, J. (1996). SUS-A quick and dirty usability scale. In I. L. Jordan, Patrick W.; Thomas, Bruce; Weerdmeester, Bernard A.; McClelland (Ed.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.

Brooke, J. (2011). *SUS - A quick and dirty usability scale (abridged)*.

Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, *8*(2), 29–40. http://www.usabilityprofessionals.org.

Brown, J. S., Collins, A. M., & Duguid, P. (2007). Situated Cognition and the Culture of Learning. *Educational Researcher*, *18*(1), 32–42.

Brown, L. M., & Brewster, S. A. (2003). Drawing by Ear: Interpreting Sonified Line Graphs. *Proceedings of the 2003 International Conference on Auditory Display*, 152–156. https://doi.org/10.1.1.597.8532

Brown, L. M., Brewster, S. A., Ramloll, R., Yu, W., & Riedel, B. (2002). Browsing Modes For Exploring Sonified Line Graphs. *Proceedings of the British Human-Computer Interface Conference*, *2*, 6–9. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.15.4378

Bruce, C. M., & Walker, B. N. (2010). Designing effective sound-based aquarium exhibit interpretation for visitors with vision impairments. *Assets (Xii)*, 251. https://doi.org/10.1145/1878803.1878856

Cahill, L. (2003). Enhanced Human Memory Consolidation With Post-Learning Stress: Interaction With the Degree of Arousal at Encoding. *Learning & Memory*, *10*(4), 270–274. https://doi.org/10.1101/lm.62403

Calderwood, C., Green, J. D., Joy-Gaba, J. A., & Moloney, J. M. (2016). Forecasting errors in student media multitasking during homework completion. *Computers and Education*, *94*(2016), 37–48. https://doi.org/10.1016/j.compedu.2015.10.021

Candey, R. M., Schertenleib, A. M., & Diaz-Merced, W. L. (2006). Xsonify sonification tool for space physics. *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, 289–290. Proceedings/2006/CandeySchertenleib2006.pdf

Charmaz, K. (2002). Qualitative Interviewing and Grounded Theory Analysis. In J. F. Gubrium & J. Holstein (Eds.), *Handbook of interview research* (pp. 675–694). Sage.

Chatzidimitris, T., Gavalas, D., & Michael, D. (2016). SoundPacman: Audio augmented reality in location-based games. *Electrotechnical Conference (MELECON), 2016 18th Mediterranean*, *April*, 1–6. https://doi.org/10.1109/MELCON.2016.7495414

Christou, C. (2010). Virtual reality in education. *Affective, Interactive and Cognitive Methods for E-Learning Design: Creating an Optimal Education Experience*, 228–243. https://doi.org/10.4018/978-1-60566-940-3.ch012

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Collins, A. M., Brown, J. S., & Newman, S. E. (1989). Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics. In L. B. Resnick (Ed.), *Knowing, Learning, and Instruction* (pp. 453–494). Lawrence Erlbaum Associates.

Corbin, J. M., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications.

D'Angelo, C., Rutstein, D., Harris, C., Haertel, G., Bernard, R., & Borokhovski, E. (2014). Simulations for STEM Learning: Systematic Review and Meta-Analysis Report Overview. In *Menlo Park: SRI International*. https://pdfs.semanticscholar.org/6bf2/15af93f56403e2fe8cb8affe798d65b1142d.pdf

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Davies, R. (2011). Understanding Technology Literacy: A Framework for Evaluating Educational Technology Integration. *TechTrends: Linking Research & Practice to Improve Learning*, *55*(5), 45–52. https://doi.org/10.1007/s11528-011-0527-3

Davison, B. K. (2013). *Understanding Graph Literacy: Understanding How Blind and Low Vision Students Can Satisfy the Core Standards with Accessible Auditory Graphs* [Georgia Institute of Technology]. https://asset1.basecamp.com/1973177/projects/3653201-8803-3-gnie/attachments/49595275/80ffcb5c80b5861a019fa19aba0297970010/original/davison_thesis.v44.pdf

Dawley, L., & Dede, C. (2014). Situated Learning in Virtual Worlds and Immersive Simulations. In *Handbook of Research on Educational Communications and Technology* (pp. 723–734). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_58

Dede, C. (2000). Emerging Influences of Information Technology on School Curriculum. *Journal of Curriculum Studies*, *32*(2), 281–303. https://pdfs.semanticscholar.org/6cbb/14465e73fe55968713d78b34106d8f084a3a.pdf

Deming, G. L. (2002). Results from the astronomy diagnostic test national project. *Astronomy Education Review*, *1*(1), 52–57. https://doi.org/10.3847/AER2001005

Desilver, D. (2017). *U.S. students' academic achievement still lags that of their peers in many other countries*. http://www.pewresearch.org/fact-tank/2017/02/15/u-s-students-internationally-math-science/

Dewey, J. (2007). *Experience and Education* (Experience). Simon and Schuster.

Diaz-Merced, W. L., Candey, R. M., Brickhouse, N., Schneps, M., Mannone, J. C., Brewster, S. A., & Kolenberg, K. (2011). Sonification of astronomical data.

Proceedings of the International Astronomical Union, *7*(S285), 133–136. https://doi.org/10.1017/S1743921312000440

Dingler, T., Lindsay, J., & Walker, B. N. (2008). Learnability of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons, and Speech. *14th International Conference on Auditory Display*, 1–6.

Dubois, M., & Vial, I. (2001). Multimedia design: the effects of relating multimodal information. *Journal of Computer Assisted Learning*, *16*(2), 157–165. https://doi.org/10.1046/j.1365-2729.2000.00127.x

Dubus, G., & Bresin, R. (2013). A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PloS ONE*, *8*(12), 1–28. https://doi.org/10.1371/journal.pone.0082491

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128.

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(1), 109–132.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the Challenges of Inquiry-Based Learning Through Technology and Curriculum Design. *Journal of the Learning Sciences*, *8*(3–4), 391–450. https://doi.org/10.1080/10508406.1999.9672075

Edwards, B. I., Bielawski, K. S., Prada, R., & Cheok, A. D. (2019). Haptic virtual reality and immersive learning for enhanced organic chemistry instruction. *Virtual Reality*, *23*(4), 363–373. https://doi.org/10.1007/s10055-018-0345-4

Ekman, I., Ermi, L., Lahti, J., Nummela, J., Lankoski, P., & Mäyrä, F. (2005). Designing sound for a pervasive mobile game. *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '05*, 110–116. https://doi.org/10.1145/1178477.1178492

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics - Physics Education Research*, *1*(1), 1–8. https://doi.org/10.1103/PhysRevSTPER.1.010103

Finn, J. D. (1993). *School Engagement and Students at Risk*.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323–327.

Finstad, K. (2013). Response to commentaries on "The usability metric for user experience." *Interacting with Computers*, *25*(4), 327–330.

https://doi.org/10.1093/iwc/iwt005

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring A New Area of Cognitive — Developmental Inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066x.34.10.906

Flowers, J. H. (2005). Thirteen Years of Reflection on Auditory Graphing: Promises, Pitfalls, and Potential New Directions. *International Conference on Auditory Displays*, 406–409.

Flowers, J. H., Buhman, D. C., & Turnage, K. D. (2005). Data sonification from the Desktop: Should Sound Be Part of Standard Data Analysis Software? *ACM Transactions on Applied Perception*, *2*(4), 467–472. https://doi.org/10.1145/1101530.1101544

Fortney, V. (2016). An Exploration of Factors that Influence Student Engagement in Science. In *ProQuest Dissertations and Theses* (Issue May). Southern Connecticut State University.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109.

Fredricks, J. A., & Mccolskey, W. (2012). Handbook of Research on Student Engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement*. Springer Science & Buusiness Media. https://doi.org/10.1007/978-1-4614-2018-7

Garcia, T., & Pintrich, P. R. (1991). Student Motivation and Self-Regulated Learning: A LISREL Model. *Annual Meeting of the American Educational Research Association*, 1–21. http://files.eric.ed.gov/fulltext/ED333006.pdf

Garcia, T., & Pintrich, P. R. (1996). Assessing Students' Motivation and Learning Strategies in the Classroom Context: The Motivated Strategies for Learning Questionnaire. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 319–339). Springer.

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, *2*, 167–177.

Gaver, W. W., Smith, R. B., & O'Shea, T. (1991). Effective sounds in complex systems: the Arkola simulation. *Proceedings of CHI*, *1991*(JANUARY 1991), 85–90. https://doi.org/10.1145/108844.108857

Georgia Department of Education. (2016). *Science Georgia Standards of Excellence*. https://www.georgiastandards.org/Georgia-Standards/Documents/Science-Eighth-Grade-Georgia-Standards.pdf

Giant Army. (2015). *Universe Sandbox* (Universe Sandbox 2). http://universesandbox.com/

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764

Gonzalez, E. T., Mitzner, T. L., Sanford, J. A., & Rogers, W. A. (2016). *TechSAge minimum battery: Overview of measures (TechSAge-TR-1601)*.

Gottfried, A. E. (1982). Relationships between academic intrinsic motivation and anxiety in children and young adolescents. *Journal of School Psychology*, *20*(3), 205–215.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology*, *77*(6), 631–645. https://doi.org/10.1037//0022-0663.77.6.631

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of Educational Psychology*, *77*, 15–46. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C11&q=greeno+collins+resnick+cognition+and+learning&btnG=&oq=Greeno+collins+re

Grond, F., & Hermann, T. (2014). Interactive Sonification for Data Exploration: How Listening Modes and Display Purposes Define Design Guidelines. *Organised Sound*, *19*(1), 41–51. https://doi.org/10.1017/S1355771813000393

Guzdial, M. (1993). *Information to users* [The University of Michigan]. https://doi.org/10.16953/deusbed.74839

Hall, R., & Nemirovsky, R. (2012). Introduction to the Special Issue: Modalities of Body Engagement in Mathematical Activity and Learning. *Journal of the Learning Sciences*, *21*(2), 207–215. https://doi.org/10.1080/10508406.2011.611447

Haller, M., Dobler, D., & Stampfl, P. (2002). Augmenting the reality with 3D sound sources. *ACM SIGGRAPH 2002 Conference Abstracts and Applications on - SIGGRAPH '02*, 65. https://doi.org/10.1145/1242073.1242100

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449–455. https://doi.org/10.1037/0003-066X.53.4.449

Harger, H., & Hyde, A. (2004). Broadcasting the Music of the Spheres: Creating Radio Astronomy. *55th International Astronautical Congress*. https://doi.org/10.2514/6.IAC-04-IAA.6.16.1.04

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction*, *17*(5), 465–477. https://doi.org/10.1016/j.learninstruc.2007.09.010

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology*, *17*(3), 300–312. https://doi.org/10.1037/0012-1649.17.3.300

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., Böcking, S., Ravaja, N., Laarni, J., Saari, T., Gouveia, F., & Sacau, A. M. (2016). The spatial presence experience scale (SPES): A short self-report measure for diverse media settings. *Journal of Media Psychology*, *28*(1), 1–15. https://doi.org/10.1027/1864-1105/a000137

Heeter, C. (1992). Being There: The Subjective Experience of Presence . *Presence: Teleoperators and Virtual Environments*, *1*(2), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

Hillman, N., & Pauletto, S. (2014). The Craftsman: the use of sound design to elicit emotions. *Intellect Limited*, *7*(1), 5–23. https://doi.org/10.1386/st.7.1.5_1

Hofer, M., Hartmann, T., Eden, A., Ratan, R., & Hahn, L. (2020). The Role of Plausibility in the Experience of Spatial Presence in Virtual Environments. *Frontiers in Virtual Reality*, *1*(April), 1–9. https://doi.org/10.3389/frvir.2020.00002

Horn, M. S., Solovey, E. T., Crouser, R. J., & Jacob, R. J. K. (2009). Comparing the use of tangible and graphical programming languages for informal science education. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 975–984. http://dl.acm.org/citation.cfm?id=1518851

Hug, D., & Misdariis, N. (2011). Towards a conceptual framework to integrate designerly and scientific sound design methods. *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, 23–30.

Inove. (2017). *Solar System Scope*. https://www.solarsystemscope.com/

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers and Education*, *106*, 150–165. https://doi.org/10.1016/j.compedu.2016.12.009

Jeon, M., Winton, R. J., Yim, J.-B., Bruce, C. M., & Walker, B. N. (2012). Aquarium fugue: interactive sonification for children and visually impaired audience in informal learning environments. *Proceedings of the 18th International Conference on Auditory Display (ICAD 2012)*, 246–247. https://smartech.gatech.edu/handle/1853/44427

Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, *8*, 7–28. https://doi.org/10.1007/BF03340893

Kaptelinin, V., & Nardi, B. (2012). Activity Theory in HCI: Fundamentals and Reflections. In *Synthesis Lectures on Human-Centered Informatics* (Vol. 5, Issue 1). https://doi.org/10.2200/S00413ED1V01Y201203HCI013

Keating, T., Barnett, M., Barab, S. A., & Hay, K. E. (2002). The virtual solar system project: developing conceptual understanding of astronomical concepts through building three-dimensional computational models. *Journal of Science Education and Technology*, *11*(3), 261–275.

Keller, C. J., Finkelstein, N. D., Perkins, K. K., & Pollock, S. J. (2006). *Assessing The Effectiveness Of A Computer Simulation In Conjunction With Tutorials In Introductory Physics In Undergraduate Physics Recitations*. 109–112.

Kemper, D. (2014). From foley to function: A pedagogical approach to sound design for novel interactions. *Journal of Sonic Studies*, *6*(1).

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220.

Kerawalla, L., Luckin, R., Seljeflot, S., & Woolard, A. (2006). "Making it real": exploring the potential of augmented reality for teaching primary school science. *Virtual Reality*, *10*(3–4), 163–174.

Kokogawa, T., Maeda, Y., Ajiki, T., Itou, J., & Munemori, J. (2012). *The effect to quality of creativity with sampling partial data from a large number of idea cards*. 147. https://doi.org/10.1145/2141512.2141564

Kramer, G. (1994). An Introduction to Auditory Displays. In G. Kramer (Ed.), *Auditory display: sonification, audification and auditory interfaces*. Addison-Wesley Publishing Company.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J. H., Miner, N., Neuhoff, J. G., Bargar, R., Barrass, S., Berger, J., Evreinov, G., Fitch, W. T., Gröhn, M., Handel, S., Kaper, H., Levkowitz, H., Lodha, S., Shinn-Cunningham, B., Simoni, M., & Tipei, S. (1999). The Sonification Report: Status of the Field and Research Agenda. In *International Community for Auditory Display (ICAD)*. http://www.researchgate.net/publication/224927615_Sonification_report_Status_of_the_field_and_research_agenda/file/32bfe510a92dcbb8c7.pdf

Kuenzi, J. J. (2008). *Science, technology, engineering, and mathematics (STEM) education: Background, federal policy, and legislative action*.

Lachlan, K., & Krcmar, M. (2011). Experiencing presence in video games: The role of presence tendencies, game experience, gender, and time spent in play.

Communication Research Reports, *28*(1), 27–31. https://doi.org/10.1080/08824096.2010.518924

Lahav, O., Kittany, J., Levy, S. T., & Furst, M. (2016). Perception of multi-varied sound patterns of sonified representations of complex systems by people who are blind. *Recent Advances on Using Virtual Reality Technologies for Rehabilitation.*, *9*(2), 139–145. http://search.proquest.com/docview/1765601916?accountid=17261%5Cnhttp://sfx.bugalicia.org/vig?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&genre=bookitem&sid=ProQ:PsycINFO&atitle=Perception+of+multi-varied+sound+patterns+of+sonified+represent

Landi, E., Alexander, R. L., Gruesbeck, J. R., Gilbert, J. A., Lepri, S. T., Manchester, W. B., & Zurbuchen, T. H. (2011). Carbon ionization stages as a diagnostic of the solar wind. *The Astrophysical Journal*, *744*(2), 1–11.

Lave, J., & Wenger, E. (1991). Legitimate Peripheral Participation. In *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (1998). A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments*, *10*(3), 282–297.

Leutner, D. (2014). Motivation and emotion as mediators in multimedia learning. *Learning and Instruction*, *29*, 174–175. https://doi.org/10.1016/j.learninstruc.2013.05.004

Levy, S. T., & Lahav, O. (2012). Enabling people who are blind to experience science inquiry learning through sound-based mediation. *Journal of Computer Assisted Learning*, *28*(6), 499–513. https://doi.org/10.1111/j.1365-2729.2011.00457.x

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5619 LNCS*, 94–103. https://doi.org/10.1007/978-3-642-02806-9_12

Lewis, J. R., Utesch, B. S., & Maher, D. E. (2013). UMUX-LITE: when there’s no time for the SUS. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2099. https://doi.org/10.1145/2470654.2481287

Liles, K. R. (2017). *Adapted Technology Experience Profile*.

Liles, K. R. (2018). Ms. An (Meeting Students’ Academic Needs): A Socially Adaptive Robot Tutor for Student Engagement in Math Education. *ProQuest Dissertations and Theses*, 105. https://search.proquest.com/docview/2070573244?accountid=49007%0Ahttp://www.yidu.edu.cn/educhina/educhina.do?artifact=&svalue=Ms.+An+%28Meeting+Students%27+Academic+Needs%29%3A+A+Socially+Adaptive+Robot+Tutor+for+Stud

ent+Engagement+in+Math+Education&stype

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality. *Educational Researcher*, *42*(8), 445–452. https://doi.org/10.3102/0013189X13511661

Liu, M., Toprac, P., & Yuen, T. T. (2008). What factors make a multimedia learning environment engaging: A case study. *Cognitive Effects of Multimedia Learning*, 173–192. https://doi.org/10.4018/978-1-60566-158-2.ch010

Liu, S. H., Liao, H. L., & Pratt, J. A. (2009). Impact of media richness and flow on e-learning technology acceptance. *Computers and Education*, *52*(3), 599–607. https://doi.org/10.1016/j.compedu.2008.11.002

Low, R., Jin, P., & Sweller, J. (2010). Learners' cognitive load when using educational technology. *Gaming and Cognition: Theories and Practice from the Learning Sciences*, 169–188. https://doi.org/10.4018/978-1-61520-717-6.ch008

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning*, *2*, 227–246.

Lunn, P., & Hunt, A. (2011). *Listening to the invisible: Sonification as a tool for astronomical discovery*. http://eprints.hud.ac.uk/15922

Ma, J., & Nickerson, J. V. (2006a). Hands-on, simulated, and remote laboratories. *ACM Computing Surveys*, *38*(3), 7-es. https://doi.org/10.1145/1132960.1132961

Ma, J., & Nickerson, J. V. (2006b). Hands-on, simulated, and remote laboratories. *ACM Computing Surveys*, *38*(3), 7-es. https://doi.org/10.1145/1132960.1132961

Mahar, M. T., Murphy, S. K., Rowe, D. A., Golden, J., Shields, A. T., & Raedeke, T. D. (2006). Effects of a classroom-based program on physical activity and on-task behavior. *Medicine and Science in Sports and Exercise*, *38*(12), 2086–2094. https://doi.org/10.1249/01.mss.0000235359.16685.a3

Mansur, D. L., Blattner, M. M., & Joy, K. I. (1985). Sound graphs: A numerical data analysis method for the blind. *Journal of Medical Systems*, *9*(3), 163–174. https://doi.org/10.1007/BF00996201

Matsumoto, D., & Sanders, M. (1988). Emotional experiences during engagement in intrinsically and extrinsically motivated tasks. *Motivation and Emotion*, *12*(4), 353–369. https://doi.org/10.1007/BF00992359

Matthews, T., Rattenbury, T., & Carter, S. (2007). Defining, designing, and evaluating peripheral displays: An analysis using activity theory. *Human-Computer Interaction*, *22*(1), 221–261.

Mayer, R. E. (2001a). A Cognitive Theory of Multimedia Learning. In *Multimedia*

Learning (pp. 41–62). Cambridge University Press.

Mayer, R. E. (2001b). *Multimedia Learning* (First). Cambridge University Press.

Mayer, R. E. (2002a). Multimedia Learning. *The Annual Report of Educational Psychology in Japan*, *41*, 27–29. https://doi.org/10.1007/978-1-4419-1428-6_285

Mayer, R. E. (2002b). Multimedia Learning. *Psychology of Learning and Motivation*, *41*, 85–139.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

McGlynn, S. A., & Rogers, W. A. (2017). *Towards a Conceptual Model of Spatial Presence in Virtual Reality*.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, *80*(4), 514–523. https://doi.org/10.1037/0022-0663.80.4.514

Miller, A.-C. S. (1990). *Effects of hands-on, activity-based science and a supportive instructional environment on at-risk sixth-grade students' attitude toward science, achievement in science, goal orientation, and cognitive engagement in science*. University of North Carolina at Greensboro.

Moore, E. B., Chamberlain, J. M., Parson, R., & Perkins, K. K. (2014). PhET interactive simulations: Transformative tools for teaching chemistry. *Journal of Chemical Education*, *91*(8), 1191–1197.

NASA. (2015). *Eyes on the Solar System*. https://eyes.jpl.nasa.gov/eyes-on-the-solar-system.html

National Governors Association. (2010). *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*.

National Governors Association Center for Best Practices, C. of C. S. S. O. (2010). *Common Core State Standards for Mathematics*. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

National Research Council. (2013). *Next generation science standards: For states, by states*. https://www.nextgenscience.org/

Neelakantan, S. (2019). *Schools Face Barriers to VR Adoption in the Classroom*. EdTech: Focus On K-12. https://edtechmagazine.com/k12/article/2019/12/schools-face-barriers-vr-adoption-classroom

Nees, M. A. (2018). Auditory Graphs Are Not the "Killer App" of Sonification, But They Work. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 1–4. https://doi.org/10.1177/1064804618773563

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In *Minnesota Studies in the PHILOSOPHY OF SCIENCE* (pp. 3–45). https://doi.org/10.1111/j.1756-8765.2009.01032.x

Nersessian, N. J. (1999). Model-based reasoning in conceptual change. In *Modelbased reasoning in scientific discovery* (Vol. 6, Issue 3, pp. 5–22). https://doi.org/10.1287/inte.19.3.58

Nielsen, J. (2012). *Usability 101: Introduction to Usability*. https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Noah, B. E., Gable, T. M., Schuett, J. H., & Walker, B. N. (2016). Forecasted Affect Towards Automated and Warning Safety Features. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - Automotive'UI 16*, 123–128. https://doi.org/10.1145/3004323.3004337

Nystrand, M., & Gamoran, A. (1991). Instructional Discourse, Student Engagement, and Literature Achievement. *Source: Research in the Teaching of English*, *25*(3), 261–290. https://doi.org/10.2307/40171413

OECD. (2019). *Programme for International Student (PISA) Results From PISA 2018 - United States*. oecd.org

Papert, S., & Harel, I. (1991). Situating Constructionism. *Constructionism*, *36*(2), 1–11. https://doi.org/10.1111/1467-9752.00269

Parker, J., & Heywood, D. (1998). The earth and beyond: developing primary teachers' understanding of basic astronomical events. *International Journal of Science Education*, *20*(5), 503–520. https://doi.org/10.1080/0950069980200501

Paterson, N., Naliuka, K., Jensen, S. K., Carrigy, T., Haahr, M., & Conway, F. (2010). Design, implementation and evaluation of audio for a location aware augmented reality game. *Proceedings of the 3rd International Conference on Fun and Games - Fun and Games '10*, 149–156. https://doi.org/10.1145/1823818.1823835

Pellerin, R., & Bouillot, N. (2009). Soundpark: Exploring ubiquitous computing through a mixed reality multi-player game experiment. *Studia Informatica …*, 1–7. http://cim.mcgill.ca/sre/publications/2009-NOTERE.pdf

Perkins, K. K., Adams, W. K., Dubson, M., Finkelstein, N. D., Reid, S., Wieman, C. E., & LeMaster, R. (2006). PhET: Interactive Simulations for Teaching and Learning Physics. *The Physics Teacher*, *44*(1), 18–23. https://doi.org/10.1119/1.2150754

PhET Interactive Simulations. (n.d.). *PhET Interactive Simulations*. http://phet.colorado.edu/

PhET Interactive Simulations. (2018). *Gravity and Oribts*. https://phet.colorado.edu/en/simulation/gravity-and-orbits%0A

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning componenets of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/http://dx.doi.org.ezproxy.lb.polyu.edu.hk/10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1991). A Manual for the User of the Motivated Strategies for Learning Questionnaire (MSLQ). In *National Center for Research to Improve Postsecondary Teching and Learning*. https://doi.org/10.5901/mjss.2015.v6n1p156

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predicitve Validity of the Motivated Strategies for Learning Questionnaire. *Educational and Psychological Measurement*, *53*, 801–813.

Pintrich, P. R., Smith, D. A. F., Gracía, T., & McKeachie, W. J. (2015). Motivated Strategies for Learning Questionnaire (MSLQ) Manual. In *National Center for Research to Improve Postsecondary Teaching and Learning*. https://doi.org/10.13140/RG.2.1.2547.6968

Podolefsky, N. S., Moore, E. B., & Perkins, K. K. (2013). Implicit scaffolding in interactive simulations: Design strategies to support multiple educational goals. *Arxciv*, 1–30. http://arxiv.org/abs/1306.6544

Post-Kammer, P., & Smith, P. L. (1985). Sex Differences in Career Self-Efficacy, Consideration, and Interests of Eighth and Ninth Graders. *Journal of Counseling Psychology*, *32*(4), 551–559. https://doi.org/10.1037/0022-0167.32.4.551

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J. S., Fretz, E., Duncan, R. G., Kyza, E., Edelson, D. C., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*(May 2014), 37–41. https://doi.org/10.1207/s15327809jls1303

Quinton, M., McGregor, I., & Benyon, D. (2016). *Sonifying the Solar System*.

Ramloll, R., Brewster, S., Yu, W., & Riedel, B. (2001). Using Non-speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users. *15th Annual Conference of the British HCI Group*, 515–529. http://eprints.gla.ac.uk/3223/

Ramloll, R., Yu, W., Brewster, S. A., Riedel, B., Burton, M., & Dimigen, G. (2000). Constructing sonified haptic line graphs for the blind student: first steps. *ACM Conference on Assistive Technologies*, 17–25.

https://doi.org/10.1145/354324.354330

Rehn, D. A., Moore, E. B., Podolefsky, N. S., & Finkelstein, N. D. (2013). Tools for high-tech tool use: A framework and heuristics for using interactive simulations. *Journal of Teaching and Learning with Technology*, *2*(1), 31–55.

Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naive Physics Reasoning: A Commitment to Substance-Based Conceptions. *Cognition and Instruction*, *18*(1), 1–34. https://doi.org/10.1207/S1532690XCI1801_01

Reisoğlu, I., Topu, B., Yılmaz, R., Karakuş Yılmaz, T., & Göktaş, Y. (2017). 3D virtual learning environments in education: a meta-review. *Asia Pacific Education Review*, *18*(1), 81–100. https://doi.org/10.1007/s12564-016-9467-0

Rogers, K. (2017). Exploring the Role of Audio in Games. *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17 Extended Abstracts*, 727–731. https://doi.org/10.1145/3130859.3133227

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Saidin, N. F., Halim, N. D. A., & Yahaya, N. (2015). A review of research on augmented reality in education: Advantages and applications. *International Education Studies*, *8*(13), 1–8. https://doi.org/10.5539/ies.v8n13p1

Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (Third). Sage Publications. https://doi.org/10.1017/CBO9781107415324.004

Sauro, J. (2015). SUPR-Q: A Comprehensive Measure of the Quality of the Website User Experience. *Journal of Usability Studies*, *10*(2), 68–86. http://www.upassoc.org.

Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P. S. (2011). Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching*, *48*(9), 1050–1078. https://doi.org/10.1002/tea.20437

Schuett, J. H. (2015). *How Many Concurrent Streams Can Listeners Utilize Within a Sonification? August*.

Schuett, J. H., & Walker, B. N. (2013). Measuring Comprehension in Sonification Tasks That Have Multiple Data Streams. *Proceedings of the 8th Audio Mostly Conference*, 1–6. https://doi.org/10.1145/2544114.2544121

Schuett, J. H., Winton, R. J., Batterman, J. M., & Walker, B. N. (2014). Auditory weather reports: demonstrating listener comprehension of five concurrent variables. *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*, 1–7.

https://doi.org/10.1145/2636879.2636898

Schweppe, J., Eitel, A., & Rummer, R. (2015). The multimedia effect and its stability over time. *Learning and Instruction*, *38*, 24–33. https://doi.org/10.1016/j.learninstruc.2015.03.001

Schweppe, J., & Rummer, R. (2016). Integrating written text and graphics as a desirable difficulty in long-term multimedia learning. *Computers in Human Behavior*, *60*, 131–137. https://doi.org/10.1016/j.chb.2016.02.035

She, H. C., & Chen, Y. Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers and Education*, *53*(4), 1297–1307. https://doi.org/10.1016/j.compedu.2009.06.012

Shelton, B. E., & Hedley, N. R. (2002). Using augmented reality for teaching Earth-Sun relationships to undergraduate geography students. *ART 2002 - 1st IEEE International Augmented Reality Toolkit Workshop, Proceedings*, 1–8. https://doi.org/10.1109/ART.2002.1106948

Shelton, B. E., & Hedley, N. R. (2004). Exploring a cognitive basis for learning spatial relationships with augmented reality. *Technology, Instruction, Cognition and Learning*, *1*(4), 323–357. http://digitalcommons.usu.edu/itls_facpub/92/

Shernoff, D. J., & Csikszentmihalyi, M. (2009). Flow in Schools: Cultivating Engaged Learners and Optimal Learning Environments DAVID. In M. J. Furlong, R. Gilman, & E. S. Huebner (Eds.), *Handbook of Positive Psychology in Schools* (pp. 131–146). Routledge. https://doi.org/10.2307/1423345

Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E. S. (2003). Student Engagement in High School Classrooms from the Perspective of Flow Theory. *School Psychology Quarterly*, *18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What It Takes to Do Well in School and Whether I ’ ve Got It : A Process Model of Perceived Control and Children ’ s Engagement and Achievement in School. *Journal of Educational Psychology*, *82*(1), 22–32.

Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *8*(5), 560–565. https://doi.org/10.1162/105474699566477

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI*, *3*(December), 1–47.

https://doi.org/10.3389/frobt.2016.00074

Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, *3*(2), 130–144.

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1007/s10750-008-9541-7

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology*, *19*(8), 1065–1087. https://doi.org/10.1002/acp.1146

Smith, D. R., & Walker, B. N. (2002). Tick-marks, Axes, and Labels: The Effects of Adding Context to Auditory Graphs. *International Conference on Auditory Display*, 1–6.

Smith, T. L., Lewis, C., & Moore, E. B. (2017). Description strategies to make an interactive science simulation accessible. *JTPD*, *5*(22), 225–238. http://scholarworks.csun.edu/handle/10211.3/190214

Smith, T. L., & Moore, E. B. (2020). Storytelling to Sensemaking: A Systematic Framework for Designing Auditory Description Display for Interactives. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3313831.3376460

Sodnik, J., Tomažič, S., Grasset, R., & Duenser, A. (2004). Spatial Sound Localization in an Augmented Reality Environment. *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, 1–8. https://doi.org/10.1145/1228175.1228197

Sperling, R. A., DuBois, N., Howard, B. C., & Staley, R. (2004). Metacognition and Self-Regulated Learning Constructs. *Educational Research and Evaluation*, *10*(2), 117–139. https://doi.org/10.1076/edre.10.2.117.27905

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Stevens, R. D., Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1994). Design and evaluation of an auditory glance at algebra for blind readers. In *Proceedings of the 2nd International Conference on Auditory Display (ICAD94)* (pp. 21–30). Proceedings/1994/StevensBrewster1994.pdf

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 55). Elsevier Inc. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tan, S.-L., Baxa, J., & Spackman, M. P. (2010). Effects of built-in audio versus unrelated background music on performance in an adventure role-playing game. *International Journal of Gaming and Computer-Mediated Simulations*, *2*(3), 1–23. https://doi.org/10.4018/jgcms.2010070101

Taylor, D. (2017). *Foley (No. 12)*. Twenty Thousand Hertz, Defacto Sound. https://www.20k.org/episodes/foley

The Collaboration for Astronomy Education Research (CAER). (1999). *Astronomy Diagnostic Test (ADT) v2*.

The Sky Live. (2018). *3D Solar System Simulator*. https://theskylive.com/3dsolarsystem

Tomlinson, B. J., Batterman, J. M., Chew, Y. C., Henry, A. G., & Walker, B. N. (2016). Exploring auditory graphing software in the classroom: The effect of auditory graphs on the classroom environment. *ACM Transactions on Accessible Computing*, *9*(1), 1–27. https://doi.org/10.1145/2994606

Tomlinson, B. J., Batterman, J. M., Kaini, P., Walker, B. N., & Moore, E. B. (2018). Supporting Simulation Use for Students with Intellectual and Developmental Disabilities. *Journal on Technology and Persons with Disabilities, 6*, 202–218.

Tomlinson, B. J., Holthausen, B. E., Edworthy, J., Hermann, T., Worrall, D., & Walker, B. N. (2019). Evaluating Auditory Displays: Perspectives from Sonification and Other Domains (Workshop). *International Conference on Auditory Display (ICAD 2019)*.

Tomlinson, B. J., Kaini, P., Harden, E. L., Walker, B. N., & Moore, E. B. (2019). A Multimodal Physics Simulation: Design and Evaluation with Diverse Learners. *Journal on Technology and Persons with Disabilities*,.

Tomlinson, B. J., Moore, E. B., & Walker, B. N. (2020). *Sound and Sonification Design for Interactive Learning Tools*. Coursera. https://www.coursera.org/learn/sound-and-sonification-for-learning

Tomlinson, B. J., Noah, B. E., & Walker, B. N. (2018). BUZZ: An Auditory Interface User Experience Scale. *CHI 2018 Extended Abstracts*, 1–6. https://doi.org/10.1145/3170427.3188659

Tomlinson, B. J., Schuett, J. H., Shortridge, W., Chandran, J., & Walker, B. N. (2016). “Talkin’’ about the weather": Incorporating TalkBack functionality and sonifications for accessible app design.” *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2016*. https://doi.org/10.1145/2935334.2935390

Tomlinson, B. J., Walker, B. N., & Moore, E. B. (2019). Using Surveys to Complete Formative Evaluations for the Design of Multimodal Interactive Simulations. *Manuscript in Preperation*.

Tomlinson, B. J., Walker, B. N., & Moore, E. B. (2020). Auditory Display in Interactive Science Simulations: Description and Sonification Support Interaction and Enhance Opportunities for Learning. *Proceedings of the 2020 CHI Conference on Human Factosrs in Computing System*, 1–12. https://doi.org/10.1145/3313831.3376886

Tomlinson, B. J., Winters, R. M., Latina, C., Bhat, S., Rane, M., & Walker, B. N. (2017). Solar System Sonification: Exploring Earth and its Neighbors Through Sound. *Proceedings of the 23rd International Conference on Auditory Display, Icad*.

Upson, R. (2002). Educational Sonification Exercises: Pathways for Mathematics and Musical Achievement. *International Conference on Auditory Display*, 1–6.

Upson, R. (2001). Sonifications as Mathematics Teaching Tools. *International Conference on Auditory Display*, 218–221.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503. https://doi.org/10.1162/105474600566989

Vazquez-Alvarez, Y., Aylett, M. P., Brewster, S. A., Jungenfeld, R. Von, & Virolainen, A. (2015). Designing Interactions with Multilevel Auditory Displays in Mobile Audio-Augmented Reality. *ACM Transactions on Computer-Human Interaction*, *23*(1), 1–30. https://doi.org/10.1145/2829944

Vezina, M. (2017). *jsOrrery*. http://mgvez.github.io/jsorrery/

Vye, N. J., Schwartz, D. L., Bransford, J. D., Barron, B. J. S., & Zech, L. (1998). SMART environments that support monitoring, reflection, and revision. In Douglas J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 305–346). Digital Printing 2009.

Vygotsky, L. S. (1978). Interaction between learning and development. In *Mind and Society* (pp. 79–91). https://doi.org/10.1016/S0006-3495(96)79572-3

Walker, B. N. (2002a). *Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification* (Vol. 8, Issue 4) [Rice]. https://doi.org/10.1037/1076-898X.8.4.211

Walker, B. N. (2002b). Magnitude Estimation of Conceptual Data Dimensions for Use in Sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211–221. https://doi.org/10.1037/1076-898X.8.4.211

Walker, B. N., Godfrey, M. T., Orlosky, J. E., Bruce, C. M., & Sanford, J. (2006). Aquarium sonification: soundscapes for accessible dynamic informal learning environments. *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006)*, 238–241. https://smartech.gatech.edu/handle/1853/50426

Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: an experimental assessment. *ACM Transactions on Applied Perception*, *2*(4), 407–412.

https://doi.org/10.1145/1101530.1101534

Walker, B. N., & Nees, M. A. (2011). Theory of Sonification. In T. Hermann, A. Hunt, & J. G. Neuhoff (Eds.), *The Sonification Handbook* (pp. 9–39). Logos Verlag.

Wang, M., Callaghan, V., Bernhardt, J., White, K., & Peña-Rios, A. (2017). Augmented reality in education and training: pedagogical approaches and illustrative case studies. *Journal of Ambient Intelligence and Humanized Computing*, *0*(0), 0. https://doi.org/10.1007/s12652-017-0547-8

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Weinburgh, M. H., & Steele, D. (2000). The Modified Attitudes toward Science Inventory: Developing an instrument to be used with fifth grade urban students. In *Journal of Women and Minorities in Science and Engineering* (Vol. 6, Issue 1, pp. 87–94). https://doi.org/10.1615/JWomenMinorScienEng.v6.i1.50

Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Sage.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Williams, D. R. (2015). *Planetary Fact Sheet*. NASA Space Science Data Coordinated Archive. http://nssdc.gsfc.nasa.gov/planetary/factsheet/

Willis, B. (1992). *Strategies for Teaching at a Distance*.

Winn, W., Windschitl, M., Fruland, R., & Lee, Y. (2002). When Does Immersion ina a virtual Environmment Helps Students construct Understanding? *Icls 2002*, *206*, 497–503. https://www.isls.org/icls/2002/

Winters, R. M., Tomlinson, B. J., Walker, B. N., & Moore, E. B. (2019). Sonic Interaction Design for Science Education. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *27*(1), 5–10. https://doi.org/10.1177/1064804618797399

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Witteman, M. J., & Segers, E. (2010). The modality effect tested in children in a user-paced multimedia environment. *Journal of Computer Assisted Learning*, *26*(2), 132–142. https://doi.org/10.1111/j.1365-2729.2009.00335.x

Wu, H.-K. K., Lee, S. W.-Y. Y., Chang, H.-Y. Y., & Liang, J.-C. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers and Education*, *62*, 41–49. https://doi.org/10.1016/j.compedu.2012.10.024

Yair, Y., Schur, Y., & Mintz, R. (2003). A "thinking journey" to the planets using scientific visualization technologies: implications to astronomy education. *Journal of Science Education and Technology*, *12*(1), 43–49. https://doi.org/10.1023/A:1022107627775

Yazzie-Mintz, E. (2007). Voices of Students on Engagement: A Report on the 2006 High School Survey of Student Engagement. *Center for Evaluation and Education Policy, Indiana University*, 12. https://files.eric.ed.gov/fulltext/ED495758.pdf%0Ahttp://ezproxy.lib.utexas.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED495758&site=ehost-live

Zeilik, M., & Morris, V. J. (2003). An Examination of Misconceptions in an Astronomy Course for Science, Mathematics, and Engineering Majors. *Astronomy Education Review*, *2*(1), 101. https://doi.org/10.3847/AER2003005

Zhao, H., Plaisant, C., & Shneiderman, B. (2005). "I hear the pattern": interactive sonification of geographical data patterns. *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems*, *2*, 1905–1908. https://doi.org/10.1145/1056808.1057052