# Evaluation Report
for

# Distance Learning Through Game-Based 3D Virtual Learning Environments: Mission Hydro Science


**Independent Evaluators**

Thomas Reeves, **University of Georgia**

William Romine, **Wright State University**

**Project Principal Investigators**

James Laffey, **University of Missouri**

Troy Sadler, **University of North Carolina**

Sean Goggins, **University of Missouri**

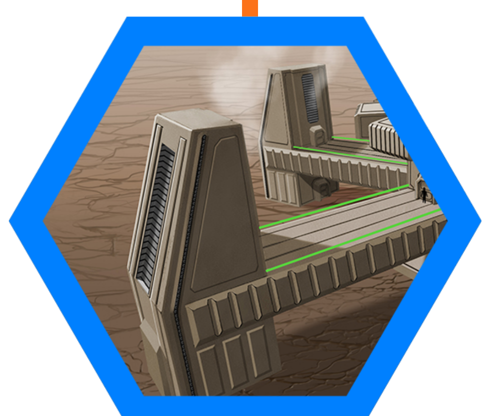

**Submitted May, 2020**


This research project was supported by grant # U411C140081 from the U.S. Department of Education's Investing in Innovation Fund.

**Contact Information:**
Thomas Reeves, **treeves@uga.edu**
William Romine, **romine.william@gmail.com**
James Laffey, **laffeyj@missouri.edu**

## Table of Contents





# Abstract

Mission HydroSci (MHS) is a 3D game-based learning environment and curriculum that supports middle school student learning of water systems science and scientific argumentation. MHS is a rigorous, coherent and engaging 6 to 8-day curriculum with all learning activities and social interactions taking place in the virtual world and with teachers observing and supporting students through an online support system enhanced by analytics. MHS was evaluated in comparison to a high-quality alternative intervention developed by the Biological Sciences Curriculum Study (BSCS) using a stratified randomized block experimental design where 'classroom' was the unit of random assignment, stratified by teacher. The comparison curriculum is called Earth's Water Systems (EWS) and is provided online using the Canvas learning management platform. Three measurable outcomes: (1) content knowledge, (2) competency in scientific argumentation, and (3) affect for science and technology were used in the pre- post-comparison of MHS with EWS. The findings of this randomized experiment showed that MHS achieved roughly equivalent water systems learning outcomes and significantly higher development of argumentation competencies when compared to the EWS curriculum. The impacts of both MHS and the EWS curriculum on affect for science and technology were equivalent and slightly negative. A secondary exploratory quasi-experimental design (QED) analysis was conducted that found significant positive effects for MHS in comparison to EWS on water systems understandings and stronger detected effects for students' argumentation.

# Part 1 - Intervention

## Description of the Intervention

Mission HydroSci (MHS) is a 3D game-based learning environment and curriculum to support middle school student learning of water systems science and scientific argumentation. The project addresses the Next Generation Science Standards (NGSS) that call for a new orientation to science teaching that prioritizes student engagement with disciplinary core ideas, crosscutting themes and scientific practices. The design of MHS is based on principles of the "transformational play" learning theory (Barab, et al 2010) which posits that students learn when they assume a character role who must use subject matter knowledge to make decisions and take action in an educational game or simulation. These actions and decisions transform the problem-based situation inherent in the game-based learning environment. The design of MHS also uses a learning progressions approach to sequencing the game play activities and

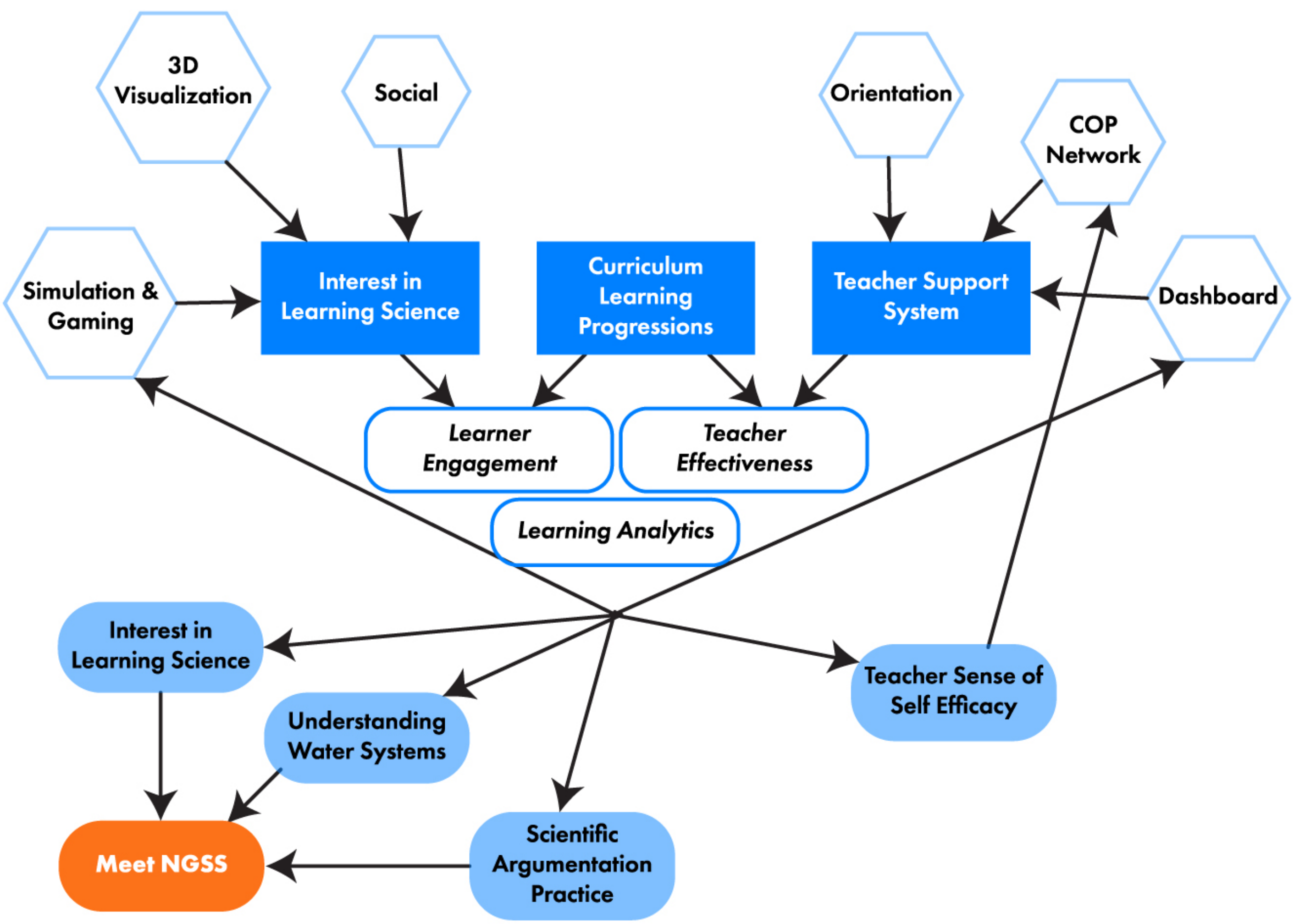



**Figure 1.** *MHS enacts learning theory and targets NGSS standards related to disciplinary core ideas, crosscutting themes and scientific practices.*

content to build upon extensive knowledge of how students make progress in learning about water systems (Covitt et al., 2009; Gunckel et al., 2009: Sadler et al., 2017) and scientific argumentation (Osborne et al. 2013).

The Logic Model (shown in figure 1) for the MHS project shows how the game play mechanisms used to enact transformational play are integrated with a learning progressions implementation of the curriculum and a way of supporting teachers that engages learning and effective teaching practices. Learning Analytics are used to provide feedback to the teachers, students and systems operation to enable continuous improvement. Thus, MHS provides a model learning system for bringing about the types of learning outcomes (disciplinary core knowledge and scientific practices) required to achieve NGSS.

MHS is a rigorous, coherent and engaging 6 to 8-day curriculum with all learning activities and social interactions taking place in the virtual world and with teachers observing and supporting students through an online support system enhanced by analytics (Laffey et al., 2017; Laffey, Griffin & Sigoloff, 2019). MHS is envisioned as a replacement curriculum component in middle school science courses addressing general and earth science. Figure 2 illustrates two screen captures from MHS.


**Figure 2.** *Non-player character (NPC) Sam tells the player about the pollution in the river, and the player tosses a sensor in river to begin the process of collecting evidence to find the pollution source.*

MHS is divided into six units that middle school students on average take 6 to 8 hours to complete.

**Unit 1** introduces students to (1) gameplay, including game controls, characters and narrative; (2) scientific argumentation as a process of using evidence to adjudicate among competing claims; and (3) the argumentation engine that will be used to build arguments during game play.

In **Unit 2** players learn about how topography influences water flow, how to use a topographic map, and what watersheds are and how the relative size of a watershed is related to the amount of water flowing through it. The player also learns to support claims with evidence.

In **Unit 3** the player must predict the spread of a dissolved material through a watershed and identify the direction of water flow based on a map of a watershed. The player also learns to identify warrants and use reasoning to link evidence with a claim.

In **Unit 4** the player learns about groundwater. Learning objectives include 1) understanding water tables, 2) predicting rates of infiltration based on permeability of the soil type, and 3) explaining the movement of water from the surface to the ground system. The player must create a complete argument (claim, reasoning and evidence).

The **Unit 5** learning objective is understanding the movement of water through a cycle, focusing on state changes that occur in atmospheric water. Students learn that energy is required for atmospheric phase changes. They also learn how to provide a counter argument to a faulty claim.

**Unit 6** is the culminating experience for players. There is a planet wide emergency unfolding and it is up to the player to figure out what is going on. It seems that water levels are dropping dramatically and if the player cannot solve the issue, the planet will no longer be viable for habitation. The player travels back to previous Unit locations and takes measurements to determine how the water levels in each have changed. In order to survive on the planet, the player must use argumentation (with a focus on the critique of arguments) to identify the cause of the problems and solve the problem of water loss.

## Description of the Comparison Curriculum

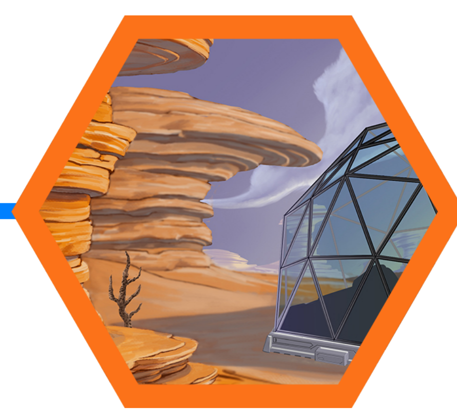
Toward evaluating the efficacy of MHS in developing NGSS type outcomes in the classroom, our evaluation compared the MHS intervention to a high-quality comparison intervention developed by the Biological Sciences Curriculum Study (BSCS). The comparison curriculum is referred to as Earth's Water Systems (EWS). The developer, BSCS, is a leader in science education curriculum development. The goal of developing EWS was to provide teachers with a high-quality alternative approach to reaching the same water systems learning objectives as were addressed in MHS.

EWS is designed to be delivered online through the Canvas learning management platform so as to help standardize the implementation across all teachers. The materials are organized in a series of lessons; the topics of these lessons and how they compare to the MHS experience are presented in table 1. Each lesson begins with a brief introduction and an opportunity for students to reveal their preexisting ideas about the lesson's content. A student progresses through an individual lesson by moving through several online pages at their own pace. Individual pages present information, explanations, and/or activities. The pages include relevant text, images, simulations, and videos.

An investigation activity is embedded within each lesson. The investigations can be performed by students in their classrooms or at home. They are demonstrations or simple experiments designed to reinforce key concepts addressed in the lesson. Toward the end of each lesson, students are invited to return to the pre-lesson assessment to reconsider the questions posed before their learning experiences and to create new answers that reflect the understandings they have constructed within the lesson. The conclusion of each lesson is a multiple-choice quiz (8-10 questions) to check student understanding. For any items that students answer incorrectly, a correct explanation is presented to the students. Once students complete a quiz and view correct answers for any questions they missed, they are able to progress to the next lesson.

| Key Water Systems Idea & Skills | MHS | EWS |
|---|---|---|
| Interpreting topographic maps | Unit 2 | Lesson 2 |
| Watersheds | Unit 2 | Lesson 2 |
| Relationship between topography & surface water | Unit 2 | Lesson 2 |
| Interpreting watershed representations | Unit 3 | Lesson 4 |
| Movement of dissolved materials in surface water | Unit 3 | Lesson 4 |
| Groundwater | Unit 4 | Lesson 3 |
| Infiltration | Unit 4 | Lesson 3 |
| Water table | Unit 4 | Lesson 3 |
| Soil types and permeability | Unit 4 | Lesson 3 |
| Water cycle | Unit 5 | Lesson 1 |
| Condensation | Unit 5 | Lesson 1 |
| Evaporation | Unit 5 | Lesson 1 |
| Precipitation | Unit 5 | Lesson 1 |

**Table 1.** *Water systems content addressed in Mission HydroSci (MHS) and Exploring Water Systems (EWS) curricula.*



# Part 2 – Study Design

## Study Sample and Setting

Thirteen middle school science teachers representing 9 schools across 6 school districts were recruited through sending notices to district science coordinators and posting a notice with the state science teachers association. All schools and teachers came from a single Midwestern state. Appendix A shows the recruitment letter sent to prospective teachers indicating their role, eligibility requirements and time frame. Eligibility required the teacher to have at least 2 class periods from 6th to 8th grade participating (one class for the comparison curriculum and at least one class for the treatment condition), conduct the comparison and treatment simultaneously in an approximate two week period between the dates of February 11 and April 15 in 2019, have suitable and available technology (Macintosh or Windows systems) for one on one computer to student instruction, be willing and able to complete the necessary training and computer setup, and to follow all protocols for supporting students. All students in a teacher's participating class at the time of random assignment participated in the study. A Child Assent and Parent Information Form was sent home with each student and there were no students or parents who objected to participation. Four of the teachers had all their classes at the 8th grade level. Seven of the teachers had all of their classes at the 7th grade level. One of the teachers had all her classes at the 6th grade level, and one of the teachers had one class at the 7th and one at the 8th grade level. An assumption was made at the beginning of the study that age differences in the students (within the ranges included in this study) would not bias the findings and that differences in prior knowledge were controlled for by having the pre-test scores as covariates in the models.

The nine schools included 4 schools with 2 participating teachers and 5 schools with a single participating teacher. All classes were considered general education classrooms and are considered blended learning situations as teachers either had computers brought to the classroom or took their students to technology laboratories so each student could be one on one with a computer. Teachers and their technology coordinators were paid a modest stipend for their participation. The 13 teachers represented public schools from both mid-sized cities and small rural communities. The student sample in the study (N=1110) included 51% male and 49% female, as well as 66% Caucasian, 11% African American, 6% Hispanic, 4% identifying as multi-racial, 3% Asian, 2% American Indian, and the remaining students self-identifying as other.

## Experimental Design

The 13 teachers participated in a stratified randomized block design where 'classroom' was the unit of random assignment, stratified by teacher. Two weeks before a teacher started the treatments one of his or her class groupings was randomly selected to undertake the comparison curriculum and the remaining classes were assigned to the MHS program. The randomization procedure was carried out by 2 research team members by assigning a number to each of the teacher's class periods and rolling a die until one of the assigned numbers came up. The class period with the die number was assigned to the comparison treatment. The study lasted 10 school days with the first and last days being used for pre and post testing. These included measures of water systems knowledge, argumentation, and affect for science and technology. All testing, pre and post testing respectively, was completed within the time of one class period (approximately 40 to 45 minutes) and on the same day for the treatment and comparison classes for each teacher. Students completed the pre and post testing using an online form with the science affect measure being given last to assure the most time for the water systems and argumentation assessments. The 13 teachers provided 48 classrooms to participate in the RCT.

Prior to random assignment, rosters of all of the study teachers' classrooms were obtained and used to identify the sample of 1,110 students. The treatment group (MHS) was comprised of 35 classes (806 students) whereas the comparison group (EWS) was comprised of 13 classes (304 students). Four students who joined treatment classrooms and two students who joined comparison classrooms after random assignment were excluded from the study. No other students received the intervention who were not included in the evaluation. Of the 1,110 students, 632 students in the MHS condition and 229 students in the comparison group completed both the pre- and post-tests for all constructs and were included in the analytic sample. There was no cluster-level attrition, and these student-level attrition rates (21.6% for MHS and 24.7% for the comparison) resulted in a total attrition rate of 22.4% and a differential attrition rate of 3.1%. The What Works Clearinghouse (WWC) considers these rates of attrition to comprise a tolerable threat of bias under conservative assumptions. Therefore, removing these incomplete cases is not likely to compromise the internal validity of the RCT.

## Instrumentation

In the comparison of a game-based curriculum with a high-quality standard curriculum, we were interested in the impact of MHS on three measurable outcomes: (1) content knowledge, (2) competency in scientific argumentation, and (3) affect for science and technology. These measures were developed or adapted from previous work by the project team. Copies of these instruments are available by contact with PI Laffey (Laffeyj@missouri.edu). Their structure and corresponding evidence for validity and reliability (Table 2) are described below.

| | N Items | Pre-Test α | Post-Test α |
|---|---|---|---|
| **Water Systems Understanding** | **23** | **0.719** | **0.815** |
| Watersheds | 6 | 0.340 | 0.452 |
| Surface Water | 3 | 0.304 | 0.503 |
| Groundwater | 4 | 0.185 | 0.341 |
| Water Cycle | 10 | 0.679 | 0.742 |
| **Argumentation Ability*** | **12** | **0.595** | **0.673** |
| Argument Alignment | 4 | 0.476 | 0.476 |
| Argument Structure | 4 | 0.434 | 0.583 |
| Argument Critique | 3 | 0.084 | 0.225 |
| **Affect for Science and Technology** | **18** | **0.906** | **0.923** |

**One item not included in subscales due to cross-loading.*

**Table 2.** *Cronbach's alpha measures for internal consistency of items within the main constructs (in* **bold***) and construct subscales.*

### Water Systems Assessment (WSA)

The WSA instrument comprises 23 multiple-choice items ($\alpha_{pre}$ = 0.719, $\alpha_{post}$ = 0.815) that address multiple dimensions of Earth water systems including watersheds (6 items, $\alpha_{pre}$ = 0.340, $\alpha_{post}$ = 0.452), surface water (3 items, $\alpha_{pre}$ = 0.304, $\alpha_{post}$ = 0.503), groundwater (4 items, $\alpha_{pre}$ = 0.185, $\alpha_{post}$ = 0.341), and water cycle processes (10 items, $\alpha_{pre}$ = 0.679, $\alpha_{post}$ = 0.742) (Sadler et al., 2017). Most of the items require application of water systems ideas (as opposed to simple recall of water facts). For example, an item related to surface water presents a watershed map and asks students to predict the movement of materials introduced to a river at a particular location. Rasch analysis also suggests adequate fit with the Rasch model (infit between 0.79 and 1.29). The 2.4 logit spread in item difficulty suggests that the items on the WSA provide information about students with a wide range of water systems knowledge (Wulff, 2019).

### Argumentation Assessment (AA)

Development of AA was informed by learning progression and assessment research related to argumentation (Osborne et al., 2013; 2014; Grooms et al., 2014). The AA is made up of 12 multiple-choice items ($\alpha_{pre}$ = 0.595, $\alpha_{post}$ = 0.673) related to a water-themed scenario. There are three item clusters that challenge students to 1) identify critical components of an argument's structure (4 items, $\alpha_{pre}$ = 0.434, $\alpha_{post}$ = 0.583), align evidence to a given claim (4 items, $\alpha_{pre}$ = 0.476, $\alpha_{post}$ = 0.476), and critique arguments (3 items, $\alpha_{pre}$ = 0.084, $\alpha_{post}$ = 0.225). One item cross-loaded onto both understanding argument structure and ability to critique arguments, and therefore was not used in the subscale calculations. Rasch model infit values ranged from 0.80 to 1.16, suggesting that students' responses are not unduly influenced by factors extraneous to their own ability and the item's difficulty. The item Rasch difficulty spread of 2.7 logits suggests that the assessment contains items suitable for measuring students at a variety of levels of argumentation (Sadler et al., 2019; Wulff, 2019).

### Student Affect

The Measure of Affect in Science and Technology (MAST) was used to measure student affect (Romine, Sadler, & Wulff, 2017). The original instrument contained 34 items that measured student interest as a main dimension and the peripheral dimensions of situational interest, attitudes towards science, interest in science careers, and interest in technology careers. We used 18 of these items ($\alpha_{pre}$ = 0.906, $\alpha_{post}$ = 0.923) focusing on use of technology in this study. These items showed adequate fit with respect to the Rasch partial credit model (infit = 0.72-1.44). Difficulty measures for the items spanned 2.4 logits and item threshold measures spanned over 5.5 logits. This provides evidence that the items used in this study yield productive measures for students at a variety of levels of affect.

# Part 3 - Analytical Approach

The analytical model was specified in line with our randomized block design stratified by teacher, where student was the Level 1 unit and classroom was the Level 2 unit. This took the form of a 2-level Hierarchical Linear Model (Raudenbush & Bryk, 2002), where students were nested within classrooms:

Level 1 (student): $$y_{ij} = \alpha_{0j} + \sum_{m=1}^{M} \alpha_{mj} X_{mij} + e_{ij},\ e_{ij} \sim N(0, \sigma^2)$$

Level 2 (classroom): $$\alpha_{0j} = \beta_{00} + \beta_{01} (MHS)_j + \mu_{0j}$$

$$\alpha_{mj} = \beta_{m0},\ m = 1, \dots, M,\ \mu_{0j} \sim N(0, \tau^2)$$

In the above system of equations, $y_{ij}$ is the outcome variable for student *i* in classroom *j*; $X_{mij}$ represents M student-level covariates including pretest, the pre-intervention covariate of water systems knowledge in the models for argumentation, and the dummy variables for the teacher blocking factor; *(MHS)* is a binary variable indicating treatment condition (*MHS* = 0 for non- MHS class; *MHS* = 1 for *MHS* class). $\beta_{01}$ represents the average impact estimate of *MHS* relative to the comparison curriculum. The significance of the impact estimate of *MHS* versus the comparison curriculum was evaluated based on the value and standard error of the coefficient $\beta_{01}$ at the 95% confidence level (2-tailed). Hedges G corrected for finite sample size (Hedges, 1981) was calculated from $\beta_{01}$ as a standardized mean difference measure for the magnitude of the impact estimate.

The effects ($\beta_{m0}$) of student-level covariates are assumed constant across classes. We included the student's teacher as a blocking variable in all of the impact analyses. The student's WSA pre-test score was included in addition to the AA pre-test score as a pre-intervention effect in the analyses of the effect of MHS on argumentation and affect for science and technology. We considered this important to include since prior content knowledge may impact a student's ability to understand and argue around water systems issues, and it is reasonable to expect that students who know more about water systems may also show greater affect for science and technology.

## Missing Data

The rates of total and differential attrition were found to provide a minimal impact on the internal validity of the RCT based on conservative assumptions outlined by the WWC. We therefore elected to not use imputation procedures, and instead conducted a complete case analysis. We elected to exclude students from the analysis who did not complete both the pre and post-tests.

# Part 4 – Results

## Confirmatory Contrasts

Results from the RCT show that MHS had small-to-negligible effects on students' understanding of water systems relative to the comparison curriculum (Table 3). The effect of MHS on knowledge of surface water systems was significant at the 2-tailed 90% confidence level ( $\beta_{01}$ =0.123, p = 0.098, Hedges G = 0.119), but the effect was small. All other effects were non-significant.

| RCT Water Systems | MHS (N = 632) | | Comparison (N= 229) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Water Systems Outcome** | **15.50** | **4.64** | **15.70** | **4.44** | **0.084** | **0.730** | **0.018** |
| **Water Systems Pretest** | **13.52** | **4.00** | **13.87** | **4.13** | | | |
| Watershed Outcome | 3.66 | 1.46 | 3.54 | 1.47 | 0.145 | 0.181 | 0.099 |
| Watershed Pretest | 2.98 | 1.39 | 3.03 | 1.47 | | | |
| Surface Water Outcome | 1.88 | 1.01 | 1.79 | 1.08 | 0.123 | 0.098 | 0.119 |
| Surface Water Pretest | 1.45 | 0.98 | 1.55 | 0.94 | | | |
| Groundwater Outcome | 2.62 | 1.05 | 2.68 | 1.04 | -0.054 | 0.470 | -0.051 |
| Groundwater Pretest | 2.28 | 1.03 | 2.32 | 1.01 | | | |
| Water Cycle Outcome | 7.34 | 2.34 | 7.69 | 2.25 | -0.216 | 0.107 | -0.093 |
| Water Cycle Pretest | 6.81 | 2.26 | 6.69 | 2.28 | | | |
| Pre-Intervention Measure: Teacher | 13 Teachers | | 13 Teachers | | | | |

**Table 3.** *Estimated effect of MHS on water systems understanding relative to the comparison curriculum based on the randomized controlled trial (RCT). Main construct in* **bold**.

Analysis of the effect of MHS on argumentation outcomes shows an impact for the game-play treatment (Table 4). MHS had a highly significant effect on the argumentation omnibus measure ( $\beta_{01}$ = 0.543, p = 0.001, Hedges G = 0.212). Analysis of the specific argumentation competencies suggests that MHS had the largest comparative effect on students' understanding of how arguments are structured ($\beta_{01}$ = 0.292, p = 0.007, Hedges G = 0.230).

We found a slight negative effect of MHS on student affect for science and technology relative to the comparison (Table 5), but this effect was not statistically significant.

| RCT Argumentation | MHS (N = 632) | | Comparison (N= 229) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Argumentation Outcome** | **7.70** | **2.58** | **7.29** | **2.47** | **0.543** | **0.001** | **0.212** |
| **Argumentation Pretest** | **6.75** | **2.43** | **6.94** | **2.35** | | | |
| Argument Alignment Outcome | 2.72 | 1.12 | 2.75 | 1.04 | 0.001 | 0.993 | 0.001 |
| Argument Alignment Pretest | 2.52 | 1.16 | 2.58 | 1.18 | | | |
| Argument Structure Outcome | 2.57 | 1.28 | 2.32 | 1.23 | 0.292 | 0.007 | 0.230 |
| Argument Structure Pretest | 2.07 | 1.22 | 2.26 | 1.12 | | | |
| Argument Critique Outcome | 1.91 | 0.83 | 1.84 | 0.77 | 0.082 | 0.158 | 0.101 |
| Argument Critique Pretest | 1.84 | 0.79 | 1.81 | 0.76 | | | |
| Pre-Intervention Measure: Teacher | 13 Teachers | | 13 Teachers | | | | |
| Pre-Intervention Measure: Water Systems | 13.52 | 4.00 | 13.87 | 4.13 | | | |

**Table 4.** *Estimated effect of MHS on argumentation relative to the comparison curriculum based on the randomized controlled trial (RCT). Main construct in* **bold**.

| RCT Affect for Science and Technology | MHS (N = 632) | | Comparison (N= 229) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Affect for Sci and Tech Outcome** | **29.43** | **10.52** | **28.95** | **10.34** | **-0.853** | **0.106** | **-0.081** |
| **Affect for Sci and Tech Pretest** | **30.91** | **9.11** | **29.84** | **9.81** | | | |
| Pre-Intervention Measure: Teacher | 13 Teachers | | 13 Teachers | | | | |
| Pre-Intervention Measure: Water Systems | 13.52 | 4.00 | 13.87 | 4.13 | | | |

**Table 5.** *Estimated effect of MHS on affect for science and technology relative to the comparison curriculum based on the randomized controlled trial (RCT). Main construct in* **bold**.

## Summary and Discussion

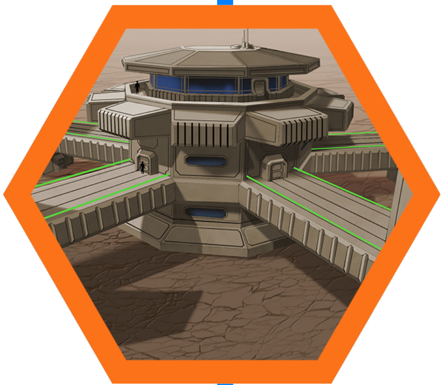
The findings of this evaluation show that Mission HydroSci (MHS) achieved roughly equivalent water systems learning outcomes and significantly higher development of argumentation competencies when compared to the high-quality comparison curriculum developed by the Biological Sciences Curriculum Study (BSCS). The impacts of both MHS and the BSCS curriculum on affect for science and technology were equivalent and slightly negative.

An important consideration in making sense of our study is that the percent completion for MHS was significantly lower than the level of completion for the BSCS comparison curriculum. The percent completion threshold required that 80% of a teacher's students reach the 4th unit of MHS (approximately half-way through the game). Only 2 of the 13 teachers met the threshold for MHS while all the EWS teachers had students reach a comparable threshold for completion. Thus, we believe the findings from the RCT to be a conservative estimate of the potential effect of MHS. When the percent completion is accounted for in an exploratory QED analysis we obtained significant positive effects of MHS on water systems understandings and stronger detected effects for students' argumentation. A more complete description of the percent completion and exploratory analyses is presented in Appendix B.

The most apparent explanation of the lower level of completion were the technological challenges to the implementation of MHS. One set of technological challenges stemmed from the quality and capabilities of the computers and computer infrastructure used to implement MHS in schools. While all the schools reported having computers that met the basic and essential requirements that we established for MHS, in fact, many of the computers did not perform to their specifications. The second set of technological challenges stem from the breadth and complexity of MHS. The project sought to bring a "high-quality" game experience to the classroom including impactful visualizations, realistic situations (such as scale of terrain when exploring a watershed), high fidelity for learning activity (such as having multiple, visualized and appropriate outcomes for decisions that players made), and implementing analytics requiring substantial data recording and processing. All of these design choices led to a complex software development project and in hindsight the final production achieved prior to the field test was not fully completed nor sufficiently tested for potential bugs across the variety of computer systems used across the different schools.

We anticipated that playing a game designed to be fun as well as educational might impact students' affect toward science. The data do not support this belief and indeed show a slight decrease in affect from pre to post testing for both MHS and the BSCS implementation. It is clear that playing MHS did not increase student interest in science or technology as measured by MAST. One explanation is that the treatment was for only a short time period relative to a middle schooler's full experience of science instruction and the use of technology in science. A second explanation might be that since the questions were not directed at the specific experience of learning science via MHS that student answers were more about prior experiences than the specific experience of playing a game in science education. A third explanation is that the implementation of MHS for the field test did not sufficiently engage the student, as we had hoped, in the role of problem solver and hero based on the use of science. Perhaps this failure can also be partly explained by the technological challenges and glitches experienced during game play, but also by our design not fully meeting our goals.

In conclusion, the evaluation results show MHS to be an effective learning experience when compared to an alternative implementation of the water systems curriculum. The MHS students achieved equivalent water systems outcomes and significantly higher development of scientific argumentation competencies. Due to the relatively low levels of progress through MHS we believe these findings to be conservative and also substantially below the outcomes we anticipated. The project team continues to work on optimization and improvement of MHS so that it can run more consistently on both lower capability computers and on the variety of configurations of computers to be found in schools. At present MHS can be a successful part of a middle school science curriculum for a school with sufficiently capable computers and technological support to overcome technical glitches. The MHS project team aims for further improvements to make MHS more broadly available to schools, teachers and students. At present, MHS also stands as a model and example of how advanced technology can be applied to achieving the goals of the NGSS, but also a cautionary story for the many challenges one can expect during the process of technology design, and carrying it through to classroom implementation.



## References


Barab, S. A., Gresfali, M. S., & Ingram-Goble, A. (2010). Transformational play using games to position person, content and context. *Educational Researcher, 39*(7), 525-536.

Covitt, B.A., Gunckel, K. L., & Anderson, C. W. (2009). Students' developing understanding of water in environmental systems. *The Journal of Environmental Education, 40*(3), 37-51.

Grooms, J., Sampson, V., & Golden, B. (2014). Comparing the effectiveness of verification and inquiry laboratories in supporting undergraduate science students in constructing arguments around socioscientific issues. *International Journal of Science Education, 36*, 1412-1433.

Gunckel, K. L., Covitt, B. A., & Anderson, C. W. (2009, June). Learning a secondary discourse: Shifts from force-dynamic to model-based reasoning in understanding water in socioecological systems. In *Learning Progressions in Science (LeaPS) Conference*, Iowa City, IA.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Laffey, J., Griffin, J., Sigoloff, J., Lander, S., Sadler, T., Goggins, S., Kim, S., Wulff, E. & Womack, A. (2017). Mission HydroSci: A progress report on a transformational role playing game for science learning. In P*roceedings of the 2017 Annual Conference of the Foundation for Digital Games*, Cape Cod, MA.

Laffey, J., Griffin, & J., Sigoloff (2019). Mission HydroSci: Gameplay to meet learning standards and more. In *Proceedings of the 2019 Annual Connected Learning Summit*, Irvine, CA.

Osborne, J., Henderson, B., MacPherson, A., & Szu, E. (2013). *Building a learning progression for argument in science*. Paper presented at the annual conference of the American Educational Research Association Conference, San Francisco, CA.

Osborne, J., Henderson, B., MacPherson, A., Wild, A., & Friend, M. (2014). *IRT analysis of items probing a unidimensional learning progression for argumentation of increasingly complex structure*. Paper presented at the annual conference of the National Association for Research in Science Teaching, Pittsburgh, PA.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.

Romine, W. L., Sadler, T. D., & Wulff, E. P. (2017). Conceptualizing student affect for science and technology at the middle school level: Development and implementation of a measure of affect in science and technology (MAST). *Journal of Science Education and Technology, 26*(5), 534-545.

Sadler, T. D., Nguyen, H., & Lankford, D. (2017). Water systems understandings: A framework for designing instruction and considering what learners know about water. *WIREs Water, 4*(1), e1178 DOI: 10.1002/wat2.1178

Sadler, T. D., Laffey, J. M., Goggins, S. P., Wulff, E. P., Womack, A. J., Griffin, J., Sigoloff, J., & Lander, S. (2019, April). *Mission HydroSci: Using gaming technologies to support NGSS-aligned learning.* NARST. Baltimore, MD.

Wulff, E. P. (2019). *Exploring the relationship between students' level of content knowledge and their ability to engage in scientific argumentation using structural equation modeling.* Unpublished dissertation. University of Missouri.

# Appendix A - Teacher Recruitment Letter

Teachers,

Thank you for your interest in the Mission HydroSci (MHS) project and your willingness to participate in our field test of our science education program. Below is a list of items that describe key features of our field test of MHS. Please review and then email Jim Laffey at LaffeyJ@missouri.edu with any questions and letting us know if you are still interested in MHS. If you are still interested, please answer the questions at the end of the email. The next step, after answering the questions, will be for us to get some information about the technology capabilities at your school for implementing MHS.

Thanks for your interest and we look forward to working with you.

+ Our MHS project includes 2 versions of a water systems curriculum for middle school students (more details about curriculum objectives at the end of this note).

+ Version A is a game students will play on windows or Mac computers and also includes support for learning scientific argumentation.

+ Version B is a set of online learning activities that you will use the learning management system CANVAS to implement with your students. (we provide CANVAS)

+ Minimum requirements for computers are to have 8 Gbytes of RAM with headsets.

+ Both versions will take 10 class periods to complete. The class periods will roughly be used as follows: 8 days of learning activity, with the first day being used for pretesting with the last day being used for post testing (more details about testing at the end of this note).

+ We will randomly assign each of your classes to version A or B.

+ You will need to have 2 classes able to participate and you will be paid a stipend for supporting our study. The stipend is to thank you for participating and in return for your completing initial training to prepare for the implementations (approximately 2-3 hours on your own schedule), distribute notices of the research to your students and their parents (if your district requires consent forms we will provide them but signed consent is not a requirement from us), implement the 10 day program, and agree to a post implementation interview and data collection.

+ The technology coordinator (or whomever at the school we will need to work with to qualify and install the computer materials for your classes) will be paid a stipend.

+ You can plan for the 2-week implementation to be anytime in the time period between February 11 and April 15.

+ The training for you will all be online and available after January 15 for you to access at your convenience.

+ Our field test procedures have been approved by the Institutional Review Board at the University of Missouri.

---

Please reply to this email letting us know of any questions you may have, your willingness to participate, and your answers to the following questions:

+ How many classes would you like to include in the study

+ Describe each class: grade level, name of course, approximate number of students, as well as any other detail you would like to share.

+ Do you have any experience using the learning management system CANVAS?

+ Do you use a learning management system or have other significant technology usage with your classes? Please describe.

+ All procedures of our study require us to comply and receive approval with the University of Missouri Institutional Review Board for protection of Human Subjects. Does your school district have a procedure for approving studies conducted with students?

---

**Both versions of the MHS curriculum are aligned with Missouri Learning Standards for middle school science. The specific learning standards addressed are:**

**Earth and Space Science**

**6-8.ESS2.C.1** Design and develop a model to describe the cycling of water through Earth's systems driven by energy from the sun and the force of gravity.

**6-8.ESS3.A** Construct a scientific explanation based on evidence for how the uneven distributions of Earth's mineral, energy, and groundwater resources are the result of past and current geoscience processes and human activity.

**6-8.ESS3.C.1** Analyze data to define the relationship for how increases in human population and per-capita consumption of natural resources impact Earth's systems.

**6-8.ESS3.C.2** Apply scientific principles to design a method for monitoring and minimizing a human impact on the environment.

**MHS is also aligned with NGSS:**

**MS-ESS2-4** Design and develop a model to describe the cycling of water through Earth's systems driven by energy from the sun and the force of gravity. [Clarification Statement: Emphasis is on the ways water changes its state as it moves through the multiple pathways of the hydrologic cycle. Examples of models can be conceptual or physical.] Assessment Boundary: A quantitative understanding of the latent heats of vaporization and fusion is not assessed.

**MHS Alignment:** Throughout the game, students are learning about the individual systems within the water cycle.

**MS-ESS3-3** Apply scientific principles to design a method for monitoring and minimizing a human impact on the environment. [Clarification Statement: Examples of the design process include examining human environmental impacts, assessing the kinds of solutions that are feasible, and designing and evaluating solutions that could reduce that impact. Examples of human impacts can include water usage (such as the withdrawal of water from streams and aquifers or the construction of dams and levees), land usage (such as urban development, agriculture, or the removal of wetlands), and pollution (such as of the air, water, or land).]

**MHS Alignment:** Within the game students are tasked with tracking the source of pollutant in watershed back to its source using scientific logic. Players are also tasked with arguing the location of the pollutant based on data collected from the environment.

**MS-ESS3-1** Construct a scientific explanation based on evidence for how the uneven distributions of Earth's mineral, energy, and groundwater resources are the result of past and current geoscience processes and human activity. [Clarification Statement: Emphasis is on how these resources are limited and typically non-renewable, and how their distributions are significantly changing as a result of removal by humans. Examples of uneven distributions of resources as a result of past processes include but are not limited to petroleum (locations of the burial of organic marine sediments and subsequent geologic traps), metal ores (locations of past volcanic and hydrothermal activity associated with subduction zones), and soil (locations of active weathering and/or deposition of rock).]

**MHS Alignment:** Within the game students are looking at the totality of the water systems on the alien planet that is featured in the game. This allows students to understand the impact of resource exploitation and its effect on a planet's water systems. In the game these effects are seen in the surface, ground and atmospheric systems.

---

Pre and Post Testing will be done with an online set of assessment instruments that include measurement of student interest in science, understanding water systems, and scientific argumentation skills.

# Appendix B - Exploratory Analysis

Prior to the study we established 8 indicators of fidelity of implementation (see Appendix C). Thresholds were met for the six indicators representing teacher activity necessary for a faithful implementation of MHS. The one student indicator for high implementation fidelity was not met. The threshold required that 80% of a teacher's students reach the 4th unit of MHS (approximately half-way through the game). Only 2 teachers met the threshold. While the threshold was simply a hypothesis about what was a meaningful dosage of game play, a contrast of student progress through MHS and the comparison curriculum shows a substantial difference in completion. The average percent completion for MHS (mean = 59.3, SD = 26.5) showed that the average student progressed 59% of the way through the MHS game and that level of progress was significantly lower than the 94% level of progress for the comparison curriculum (mean = 94.5, SD = 17.0). This difference in progress was significant at the 95% confidence level (Mann-Whitney U = 7879.5, $p \ll 0.001$). Two teachers only had 13% and 28% of students reaching half-way through MHS and 4 teachers only had approximately 50% of students reaching half-way.

Due to the relatively low levels of student progress through MHS we believe the effects detected in the RCT can be considered highly conservative and likely to underestimate the true effect. As an exploratory analysis, we conducted a quasi-experimental design (QED) that breaks the random assignment in order to provide impact estimates adjusted for percent completion. We did this in two ways: first by eliminating classrooms for the 2 teachers who did not meet fidelity thresholds for student completion of MHS (13% and 28% of the students in the classrooms for these teachers only made it at least to Level 4 in the game – approximately midway through the game), and second by keeping all of the classrooms in the analysis and adjusting for percent completion.

For the first QED analysis we removed the two lowest-fidelity teachers from the analysis, leaving 40 classrooms over 11 teachers. Since 'teacher' was the blocking variable for the stratified random assignment, removing these two teachers also constituted removing their comparison classrooms. This yielded 534 complete cases for the MHS curriculum and 199 for the comparison. The incomplete cases were excluded from the analysis. For a more liberal estimate of the intervention effect a second exploratory analysis was conducted by adjusting the effects for percent completion of the MHS and comparison curricula, respectively. Scores for percent completion for MHS were derived from the game log based on the level completed in the game. Scores for percent completion of the comparison curriculum were calculated based on progress through course modules. Percent completion was included within the model as a covariate, thereby adjusting the intervention effect for the amount of the curricula completed. Among the full sample of students from the RCT design, we were able to match logs to the test scores for 572 out of the 632 complete cases for MHS and 218 out of the 229 complete cases for the comparison curriculum. The rest of the cases were excluded from the analysis. Baseline equivalence was calculated as the standardized mean difference between the raw unadjusted pre-test means of the MHS and comparison groups.

## Exploratory Contrasts

When the two low-fidelity teachers are removed (Tables B1-3), the conclusion regarding the effect of MHS on knowledge of water systems and argumentation remains the same as that derived from the RCT (Tables 3 and 4). Understanding of surface water systems (Table B1) increased slightly, but the effect was still small and significant at the 2-tailed 90% confidence level ($\beta_{01}$ =0.145, p = 0.080, Hedges G = 0.140). The effects of MHS on students' argumentation ($\beta_{01}$ =0.662, p < 0.001, Hedges G = 0.256) and understanding of the structure of arguments ($\beta_{01}$ =0.378, p = 0.001, Hedges G = 0.297) (Table B2) also increased slightly over those found in the RCT, but nonetheless bear a similar qualitative interpretation. All other effects remained non-significant.

| QED Water Systems | MHS (N = 534) | | Comparison (N= 199) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Water Systems Outcome** | **15.72** | **4.62** | **15.70** | **4.48** | **0.199** | **0.436** | **0.043** |
| **Water Systems Pretest** | **13.60** | **3.99** | **13.82** | **4.21** | | | **-0.054*** |
| Watershed Outcome | 3.70 | 1.43 | 3.55 | 1.49 | 0.164 | 0.136 | 0.113 |
| Watershed Pretest | 2.99 | 1.40 | 3.04 | 1.49 | | | -0.035* |
| Surface Water Outcome | 1.91 | 1.02 | 1.80 | 1.08 | 0.145 | 0.080 | 0.140 |
| Surface Water Pretest | 1.46 | 0.97 | 1.57 | 0.97 | | | -0.113* |
| Groundwater Outcome | 2.70 | 1.04 | 2.67 | 1.04 | 0.015 | 0.852 | 0.014 |
| Groundwater Pretest | 2.32 | 1.03 | 2.33 | 1.01 | | | -0.010* |
| Water Cycle Outcome | 7.41 | 2.32 | 7.67 | 2.26 | -0.184 | 0.204 | -0.080 |
| Water Cycle Pretest | 6.84 | 2.25 | 6.89 | 2.32 | | | -0.022* |
| Pre-Intervention Measure: Teacher | 11 Teachers | | 11 Teachers | | | | |

*Satisfies baseline equivalence. Pre-test is always included in the model to adjust for baseline differences.

**Table B1.** *Estimated effect of MHS on water systems understanding relative to the comparison curriculum based on a quasi-experimental design (QED) which removes two low-fidelity teachers. Main construct in* **bold**.

The fact that removing the two low-fidelity teachers improved the magnitude of the effects suggests that an adjustment for amount of the curricula completed may yield a more realistic estimate of the effect one might expect to obtain for a group of students who complete the entire game relative to a group completing the entire comparison curriculum. After this correction for percent completion was applied (Tables B4-6), we obtained significant positive effects of MHS on water systems understandings ( $\beta_{01}$ =0.813, p = 0.007, Hedges G = 0.177) (Table B4). This significant effect was primarily due to highly significant gains in knowledge of watersheds ( $\beta_{01}$ =0.480, p < 0.001, Hedges G = 0.325) and surface water systems ( $\beta_{01}$ =0.360, p < 0.001, Hedges G = 0.246). The conclusions for argumentation and affect (Tables B5 and B6) were similar to the RCT and the previous QED analysis, only with stronger detected effects for students' argumentation ( $\beta_{01}$ =0.888, p < 0.001, Hedges G = 0.348) and understanding of how arguments are structured ( $\beta_{01}$ =0.493, p < 0.001, Hedges G = 0.388).

| QED Argumentation | MHS (N = 534) | | Comparison (N= 199) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Argumentation Outcome** | **7.78** | **2.60** | **7.20** | **2.53** | **0.662** | **0.000** | **0.256** |
| **Argumentation Pretest** | **6.75** | **2.44** | **6.81** | **2.33** | | | **-0.025*** |
| Argument Alignment Outcome | 2.72 | 1.13 | 2.70 | 1.07 | 0.039 | 0.626 | 0.035 |
| Argument Alighnment Pretest | 2.51 | 1.17 | 2.53 | 1.21 | | | -0.017* |
| Argument Structure Outcome | 2.61 | 1.28 | 2.27 | 1.25 | 0.378 | 0.001 | 0.297 |
| Argument Structure Pretest | 2.07 | 1.24 | 2.21 | 1.09 | | | -0.116* |
| Argument Critique Outcome | 1.93 | 0.82 | 1.86 | 0.79 | 0.075 | 0.254 | 0.092 |
| Argument Critique Pretest | 1.84 | 0.80 | 1.81 | 0.77 | | | 0.038* |
| Pre-Intervention Measure: Teacher | 11 Teachers | | 11 Teachers | | | | |
| Pre-Intervention Measure: Water Content | 13.60 | 3.99 | 13.82 | 4.21 | | | |

*Satisfies baseline equivalence. Pre-test is always included in the model to adjust for baseline differences.

**Table B2.** *Estimated effect of MHS on argumentation relative to the comparison curriculum based on a quasi-experimental design (QED) which removes two low-fidelity teachers. Main construct in* **bold**.

| QED Affect for Science and Technology | MHS (N = 534) | | Comparison (N= 199) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Affect for Sci and Tech Outcome** | **29.21** | **10.52** | **28.35** | **10.36** | **-0.804** | **0.159** | **-0.077** |
| **Affect for Sci and Tech Pretest** | **30.47** | **9.16** | **29.12** | **9.90** | | | **0.144*** |
| Pre-Intervention Measure: Teacher | 11 Teachers | | 11 Teachers | | | | |
| Pre-Intervention Measure: Water Systems | 13.60 | 3.99 | 13.82 | 4.21 | | | |

*Satisfies baseline equivalence. Pre-test is always included in the model to adjust for baseline differences.

**Table B3.** *Estimated effect of MHS on affect for science and technology relative to the comparison curriculum based on a quasi-experimental design (QED) which removes two low-fidelity teachers. Main construct in* **bold**.

| QED Water Systems | MHS (N = 572) | | Comparison (N= 218) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Water Systems Outcome** | **15.41** | **4.63** | **15.66** | **4.50** | **0.813** | **0.007** | **0.177** |
| **Water Systems Pretest** | **13.42** | **3.98** | **13.84** | **4.11** | | | **0.104***|
| Watershed Outcome | 3.63 | 1.46 | 3.54 | 1.50 | 0.480 | 0.000 | 0.325 |
| Watershed Pretest | 2.94 | 1.38 | 3.03 | 1.47 | | | 0.064* |
| Surface Water Outcome | 1.85 | 1.02 | 1.78 | 1.09 | 0.360 | 0.000 | 0.346 |
| Surface Water Pretest | 1.43 | 0.97 | 1.54 | 0.94 | | | 0.114* |
| Groundwater Outcome* | 2.61 | 1.03 | 2.67 | 1.03 | 0.133 | 0.160 | 0.129 |
| Groundwater Pretest* | 2.27 | 1.03 | 2.33 | 1.01 | | | 0.059* |
| Water Cycle Outcome | 7.32 | 2.36 | 7.67 | 2.28 | 0.126 | 0.462 | 0.054 |
| Water Cycle Pretest | 6.78 | 2.27 | 6.94 | 2.29 | | | 0.070* |
| Percent Completion | 59.34 | 26.50 | 94.50 | 17.03 | | | |
| Pre-Intervention Measure: Teacher | 13 Teachers | | 13 Teachers | | | | |

*Satisfies baseline equivalence. Pre-test is always included in the model to adjust for baseline differences.
**Table B4.** *Estimated effect of MHS on water systems understanding relative to the comparison curriculum based on a quasi-experimental design (QED) where measures of impact are adjusted for percent completion. Main construct in* **bold**.

| QED Argumentation | MHS (N = 572) | | Comparison (N= 218) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Argumentation Outcome** | **7.63** | **2.57** | **7.26** | **2.49** | **0.888** | **0.000** | **0.348** |
| **Argumentation Pretest** | **6.70** | **2.42** | **6.92** | **2.38** | | | **0.091 *** |
| Argument Alignment Outcome | 2.69 | 1.13 | 2.75 | 1.05 | 0.063 | 0.511 | 0.057 |
| Argument Alighnment Pretest | 2.50 | 1.17 | 2.57 | 1.20 | | | 0.059* |
| Argument Structure Outcome | 2.53 | 1.28 | 2.30 | 1.23 | 0.493 | 0.000 | 0.388 |
| Argument Structure Pretest | 2.06 | 1.22 | 2.26 | 1.13 | | | 0.167* |
| Argument Critique Outcome | 1.90 | 0.83 | 1.84 | 0.78 | 0.105 | 0.157 | 0.129 |
| Argument Critique Pretest | 1.82 | 0.80 | 1.81 | 0.76 | | | -0.013* |
| Percent Completion | 59.34 | 26.50 | 94.50 | 17.03 | | | |
| Pre-Intervention Measure: Teacher | 13 Teachers | | 13 Teachers | | | | |
| Pre-Intervention Measure: Water Content | 13.42 | 3.98 | 13.84 | 4.11 | | | |

*Satisfies baseline equivalence. Pre-test is always included in the model to adjust for baseline differences.
**Table B5.** *Estimated effect of MHS on argumentation relative to the comparison curriculum based on a quasi-experimental design (QED) where measures of impact are adjusted for percent completion. Main construct in* **bold**.

| QED Affect for Science and Technology | MHS (N = 572) | | Comparison (N= 218) | | | | |
|---|---|---|---|---|---|---|---|
| Measure | Raw Mean | SD | Raw Mean | SD | Impact Est. | p-value | Effect Size |
| **Affect for Sci and Tech Outcome** | **29.14** | **10.63** | **29.02** | **10.23** | **-0.048** | **0.943** | **-0.005** |
| **Affect for Sci and Tech Pretest** | **30.77** | **9.23** | **29.97** | **9.74** | | | **-0.085*** |
| Percent Completion | 59.34 | 26.50 | 94.50 | 17.03 | | | |
| Pre-Intervention Measure: Teacher | 13 Teachers | | 13 Teachers | | | | |
| Pre-Intervention Measure: Water Systems | 13.42 | 3.98 | 13.84 | 4.11 | | | |

*Satisfies baseline equivalence. Pre-test is always included in the model to adjust for baseline differences.

**Table B6.** *Estimated effect of MHS on affect for science and technology relative to the comparison curriculum based on a quasi-experimental design (QED) where measures of impact are adjusted for percent completion. Main construct in* **bold**.

# Appendix C - Mission HydroSci Fidelity Report

In the Dev89_DesignSummary_10-10-2018 Evaluation plan (MHS_fidelity table_10-10-18) approved by the Abt Associates Analysis and Reporting Team, the MHS project committed to assessing fidelity of implementation of the 8 key components with a measure for each component. Figure C1 shows the logic model for the evaluation plan identifying 8 components necessary for fidelity.

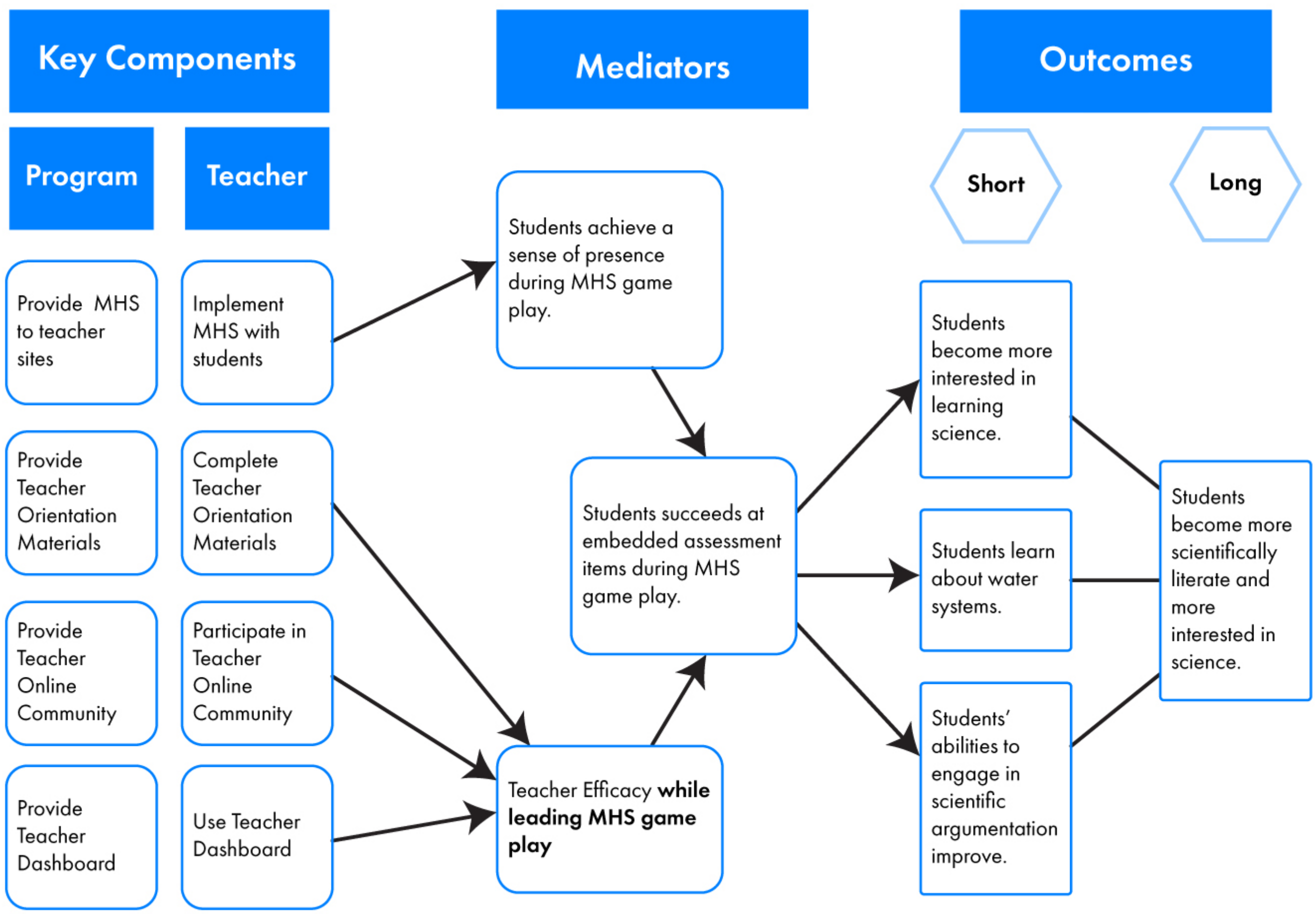



**Figure C1.** *The logic model for the evaluation plan identifying 8 components necessary for fidelity.*

We noted in the Evaluation Plan that while the standards for i3 projects call for 2 years of fidelity data, our project would only collect one year of data. Because of the substantial design and development work to complete the 3D game-based virtual learning system, the full system was only able to be tested in year 5 (final year) of the project. Sign off was granted by OII for including only one year of data for fidelity reporting.

The eight fidelity assessments are identified in the MHS_fidelity table_10-10-18 document and the

| Intervention Component | Implementation Measure | Number Of Units In Which Fidelity Components Was Measured | Number Of Units In Which The Intervention Was Implemented | Component Level Threshold For Fidelity Of Implementation For The Unit That Is The Basis For The Sample Level | Evaluator s Criteria For "Implemented With Fidelity" At Sample Level | Component Level Fidelity Score For The Entire Sample | Implemented With Fidelity? (Yes, No, N/A) |
|---|---|---|---|---|---|---|---|
| Planned Intervention Activities [i.e., key components] | | | | | | | |
| Provide MHS to teacher sites | 1 | 13 teachers | 13 teachers | Program provides teachers with MHS curriculum and all materials<br>Score 2 out of 2 | At least 80% of teachers have a score of 2 | 13 teachers met threshold | Yes |
| Implement MHS with students | 1 | 13 teachers, 572 students | 13 teachers, 632 students | Students are able to play the learning game. High implementing student = score of 4 or more out of 6. High implementing teacher = 80% of students with score of 4 or more | At least 80% of teachers by classrooms are high implementing | 2 teachers met threshold | No |
| Provide Teacher Orientation Materials (TOM) | 1 | 13 teachers | 13 teachers | Program provides teachers with materials to prepare for teaching MHS.<br>Score 1 out of 1 | At least 80% of teachers have a score of 1 | 13 teachers met threshold | Yes |
| Complete Teacher Orientation Materials | 1 | 13 teacher | 13 teachers | Teacher fulfills expectations for Undertaking preparation for teaching MHS.<br>Score 1 (or higher) out of 2 | At least 80% of teachers have a score of 1 or higher | 13 teachers met threshold | Yes |
| Provide Teacher Online Community (TOC) | 1 | 13 teachers | 13 teachers | Program provides teachers with MHS support materials<br>Score 1 out of 1 | At least 80% of teachers have a score of 1 | 13 teachers met threshold | Yes |
| Participate in Teacher Online Community | 1 | 13 teachers | 13 teachers | Teacher fulfills expectations for using MHS support materials.<br>Score 1 (or higher) out of 2 | At least 80% of teachers have a score of 1 or higher | 13 teachers met threshold | Yes |
| Provide Teacher Dashboard | 1 | 13 teachers | 13 teachers | Program provides teachers with MHS dashboard for each of their classes.<br>Score 1 out of 1 | At least 80% of teachers have a score of 1 | 13 teachers met threshold | Yes |
| Use Teacher Dashboard | 1 | 13 teachers | 13 teachers | Teacher fulfills expectations for using performance support for teaching MHS.<br>Score 1 (or higher) out of 2 | At least 80% of teachers have a score of 1 or higher | 13 teachers met threshold | Yes |

**Table C1.** *Results of Fidelity assessments for the field test of MHS*

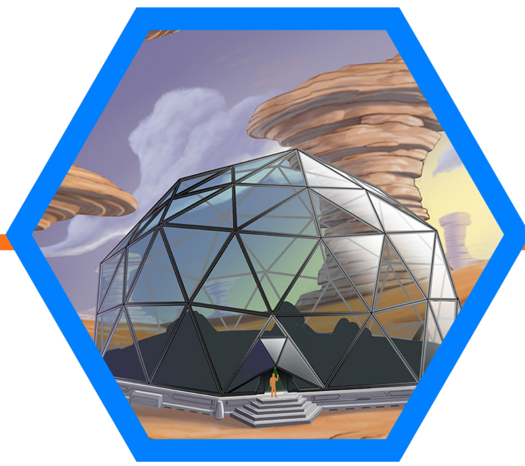
**Table C2.** *The original fidelity table approved for the study by the Abt Associates Analysis and Reporting Team.*

| Component | Definition | Criteria | Measure | Item Metric | Component Unit Threshold (1) | Component Unit Threshold (2) | Program Threshold | Sample To Be Measured | Years Of Measurement |
|---|---|---|---|---|---|---|---|---|---|
| Provide MHS to teacher sites | Program provides teachers with MHS curriculum and all materials | Teacher logs into MHS approx. one week prior to GameStart and completes startup_test protocol. (applies to all MHS classes taught by the teacher) | 1. MHS provides data to MHS_datacentral showing completion of startup_test protocol.<br>2. Teacher responds to email inquiry stating successful completion of startup_test protocol. | Teacher implementation score ranges from 0 – 2<br>0=no data or email response on startup test protocol completion<br>1=data or email response on startup test protocol completion<br>2= data and email response on startup test protocol completion | | High implementing teacher = score of 2 | High fidelity at the program level = at least 80% of teachers have a score of 2 | All teachers in the study sample | Year 5 of the study (1 year impact study) |
| Implement MHS with students | Students are able to play the learning game. | Student logs into MHS units 1-6. | MHS provides data to MHS_datacentral showing student login for each unit (1-6). | Student implementation score ranges from 0 – 6 (number of unit logins) | High implementing student = score of 4 or more | High implementing teacher = 80% of students with score of 4 or more (noted for each class of each teacher) | High fidelity at the program level = at least 80% of teachers by classrooms are high implementing | All students and teachers in the study sample | Year 5 of the study (1 year impact study) |
| Provide Teacher Orientation Materials (TOM) | Program provides teachers with materials to prepare for teaching MHS. | Teacher logs into MHS teacher support (MHSts) site and accesses orientation materials. (applies to all MHS classes taught by the teacher) | MHS provides data to MHS_datacentral showing teacher access. | Teacher implementation score ranges from 0 – 1<br>0 = teacher does not login to access orientation materials<br>1 = teacher logs in to access orientation materials | | High implementing teacher = score of 1 | High fidelity at the program level = at least 80% of teachers have a score of 1 | All teachers in the study sample | Year 5 of the study (1 year impact study) |
| Complete Teacher Orientation Materials | Teacher fulfills expectations for undertaking preparation for teaching MHS. | Teacher completes orientation checklist on MHSts meeting minimum criteria. (applies to all MHS classes taught by the teacher) | Checklist is evaluated using a rubric where 1 is a minimum criteria on a scale of 0-2 | Teacher implementation score ranges from 0 -2 | | High implementing teacher = score of 1 or higher | High fidelity at the program level = at least 80% of teachers have a score of 1 or higher | All teachers in the study sample | Year 5 of the study (1 year impact study) |

| Component | Definition | Criteria | Measure | Item Metric | Component Unit Threshold (1) | Component Unit Threshold (2) | Program Threshold | Sample To Be Measured | Years Of Measurement |
|---|---|---|---|---|---|---|---|---|---|
| Provide Teacher Online Community (TOC) | Program provides teachers with MHS support materials | Teacher logs into MHS teacher site and accesses Online Community (MHSoc). (applies to all MHS classes taught by the teacher) | MHSoc provides data to MHS_ datacentral showing teacher access. | Teacher implementation score ranges from 0 – 1<br>0 = teacher does not login to access online community<br>1 = teacher logs in to access online community | | High implementing teacher = score of 1 | High fidelity at the program level = at least 80% of teachers have a score of 1 | All teachers in the study sample | Year 5 of the study (1 year impact study) |
| Participate in Teacher Online Community | Teacher fulfills expectations for using MHS support materials. | Teacher completes Online Community evaluation form on MHSts meeting minimum criteria. (applies to all MHS classes taught by the teacher) | Evaluation form is evaluated using a rubric where 1 is a minimum criteria on a scale of 0-2 | Teacher implementation score ranges from 0 -2 | | High implementing teacher = score of 1 or higher | High fidelity at the program level = at least 80% of teachers have a score of 1 or higher | All teachers in the study sample | Year 5 of the study (1 year impact study) |
| Use Teacher Dashboard | Teacher fulfills expectations for using performance support for teaching MHS. | Teacher accesses Teacher Dashboard for each unit (1-6) meeting minimum criteria. | MHS provides data to MHS_datacentral showing Dashboard use and use pattern is evaluated using a rubric where 1 is a minimum criteria on a scale of 0-2 | Teacher implementation score ranges from 0 - 2 | | High implementing teacher = score of 1 or higher | High fidelity at the program level = at least 80% of teachers have a score of 1 or higher | All teachers in the study sample | Year 5 of the study (1 year impact study) |

fidelity results are summarized below in Table C1. The sample sizes were 13 teachers and 572 students. The data collection included 632 students participating in MHS but we have progress records only for 572 students. Progress records for the other 60 are not available because of data loss in the dashboard mechanism for capturing progress. Table C2 presents the original fidelity table approved for the study by the Abt Associates Analysis and Reporting Team showing more information about the measurement of the fidelity components.

The work described herein is supported by the US Department of Education's Investing in Innovation (i3) program (U411C140081). The ideas expressed are those of our project team and do not necessarily reflect the views of the funders.




Mission hydrosci