TABLE OF CONTENTS:

A. SIGNIFICANCE	2
…. Absolute Priority	2
…. Novel Approach	3
…. Learner Experience	5
…. Contribution to Theory, Knowledge and Practice	6
B. QUALITY OF DESIGN	6
…. Goals	6
…. Logic Model	7
…. Activities	8
…. Output	12
…. Potential Risks	15
C. MANAGEMENT PLAN & PERSONNEL	16
…. PERSONNEL	18
E. PROJECT EVALUATION	21
…. Impact Study Research Questions	21
…. Measurable Thresholds for Implementation	25
…. Resources for Carrying Out Evaluation	26

A. SIGNIFICANCE - The proposed project addresses ***Priority 5 Effective Use of Technology*** *by (subpart b) integrating technology with the implementation of rigorous standards to increase student achievement and engagement and teacher efficacy.*

The Mission HydroScience (MHS) team seeks a development award to design, develop and evaluate a game-based 3D virtual learning environment (3D VLE) for teaching and learning in blended or distance education. MHS targets middle school students learning hydrologic systems and scientific argumentation. The *Next Generation Science Standards* (NGSS) call for a new orientation to science teaching and learning that prioritizes student engagement with disciplinary core ideas, crosscutting themes and scientific practices. Implementation of NGSS is an ambitious challenge even in well-resourced classrooms, but as a nation we must also attend to students with diverse learning needs in small and rural communities. Online learning, which includes students learning at a distance and in blended environments, is an approach to providing high quality instruction to diverse students in diverse settings. However, innovation is needed before online learning can readily deliver the deep engagement with core ideas, crosscutting themes and scientific practices envisioned for NGSS.

Small and rural schools, in particular, are turning to online distance learning as a mechanism for addressing the challenges of attracting and keeping effective teachers and expanding course selections for their students (Hannum et al., 2009). Distance and blended learning models can increase the number of students participating in high quality science education. However, prevalent forms of online learning have attrition rates sometimes exceeding 50%. Students report isolation, frustration and lack of support from traditional information-centric online teaching (Roblyer 2006 & Varre et al. 2011). Prevalent online learning offerings are not consistent with what we know about how people learn (Bransford 1999).

Video game play is becoming ubiquitous in American households and places the game player in an active, deeply engaging context where actions and decisions have consequences for progression in the game. Game play also encourages learning from failure, perseverance and sense of identity. Techniques for virtualization and game playing experiences can also be applied to 3D virtual learning. Projects such as iSocial (Laffey, Stichter & Galyen, 2013a; 2013b), *River City* (Clarke et al., 2006), *Mission Biotech* (Sadler et al. 2013), *Quest Atlantis* (Barab, Sadler et al. 2007), *EcoMUVE* (Metcalf et al. 2009) and *SimCityEDU* (Glasslab, 2013) demonstrate that 3D VLE engages students and can produce significant student outcomes. A recent review (NRC, 2011) of the role of games and simulations in science education suggests that these technologies may be an important approach for achieving NGSS aligned learning. The report concludes that simulations can be effective in developing conceptual understanding, while the evidence for games is still emerging. We propose an iterative, design process consistent with research type 3 of the IES/NSF Common Guidelines with methods for developing an intervention and collecting evidence of feasibility and outcomes along with external critical review.

Novel Approach. We will use "strong theory" as a basis for learning progressions through levels of understanding water systems and argumentation competencies (Osborne et al. 2013), and the theory of transformational play as a method for integrating simulation and game play for learning. We will use technology to scale highly effective teaching and learning practices to meet diverse learning needs. Transformational play (Barab, et al 2010) includes the student taking a role (playing a protagonist) who must use subject matter knowledge to make decisions and take action during play. These actions and decisions transform the problem-based situation. In turn, the student's understanding of the subject matter and identity is transformed through the process of game play. Our vision for a simulation environment for hydroscience and a game-based drama

with non-player characters for developing argumentation is unique from 3D virtual learning efforts. Among a variety of distinctions, three are noted here. First we plan to employ Learning Analytics to create an adaptive system for student learning and assessment and provide monitoring and awareness for teachers. We will develop analytics that focus on tracking individuals' specific choices, then analyzing those discrete choices against a backdrop of learning outcomes and argumentation competencies; **thus assessment is built into playing the game.** Second, current systems emphasize individuals performing certain tasks somewhat piecemeal in a virtual world while interacting with peers and the instructor in the physical classroom for a relatively short period. Typically, teachers monitor student work by walking around the class and observing physical behavior. In contrast, MHS will be a rigorous, coherent and engaging 4-week curriculum with all learning activities and social interactions taking place in the virtual world and with teachers observing and supporting students through analytics. This configuration is part of our design because for students at a distance the student-teacher relationship is mediated fully through the VLE. Third, teachers need support to be effective in teaching in online environments. In addition to traditional teacher support materials helping orient and providing practice for the new elements of the teaching role, we plan a networked community of practice and a dashboard to visualize student activity and progress. The community of practice will be built through an online site supporting interaction with other MHS teachers and MHS project personnel to support preparation and problem solving. The dashboard for visualizing student activity will be designed from a performance support framework to optimize acting upon insights such as recognizing when a student is falling behind and adding an additional support to the next lesson to help structure the activity for the student.

Learner Experience: MHS will be developed in the 3D game engine Unity (unity3D.com), and will include two components: 1) a domain knowledge curriculum and 2) a scientific argumentation environment (SA). The interface between the domain curriculum and SA will be data cards (drawn from simulations or experiments) and information cards (drawn from learning materials such as information the student may look up in a virtual library). The game implementation confronts the student with a futuristic world in which the Earth's natural resources are overused. The student must colonize other planets where fresh water is potentially available and can be managed to sustain a population. We envision a multi-level game wherein early experiences provide opportunities for players to engage in tasks such as creating and mapping watersheds. As play progresses, the game challenges learners to make the case for a solution to populating a landmass with a human community taking into account the requisite water needs. In the game, the learner takes on the identity of a scientist conducting experiments while testing and building the case for solutions.

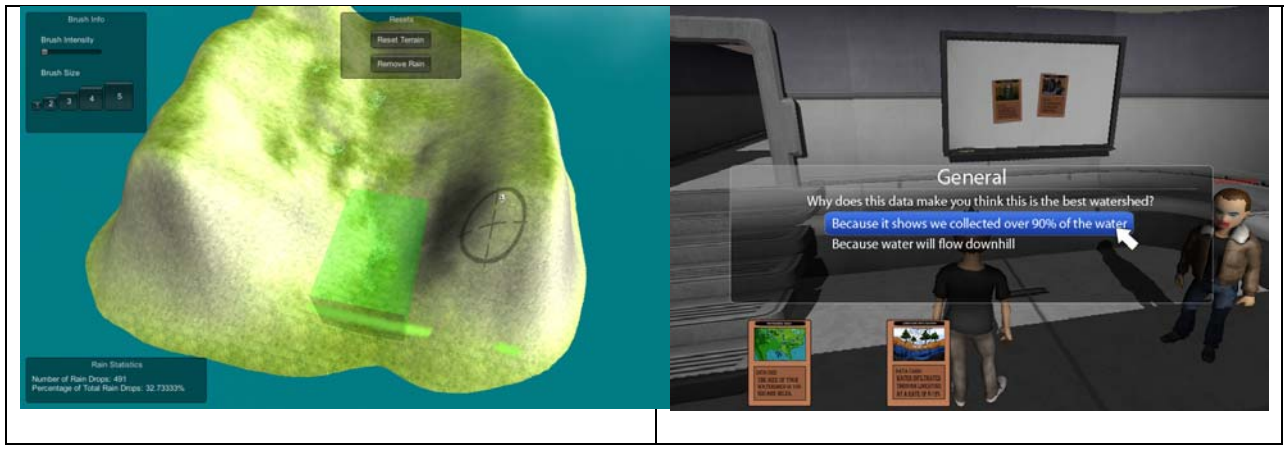


Figure 1. Screenshots from preliminary MHS work. On the left the student is manipulating a landmass to create a watershed. On the right the student is defending claims to a supervisor.

The learner is supported by non-player characters (NPCs) who can guide and assist the work; the learner also confronts NPC antagonists who challenge assertions made by the learner and her/his collaborators and in some instances introduce false claims to the discourse. At each

level of the game the learner is given missions to explore within the domain knowledge experience such as creating a watershed, developing data and information cards from the water systems experience, and engaging in discussion with NPCs (and in later levels with peers) to both explain and use the cards in argumentation events.

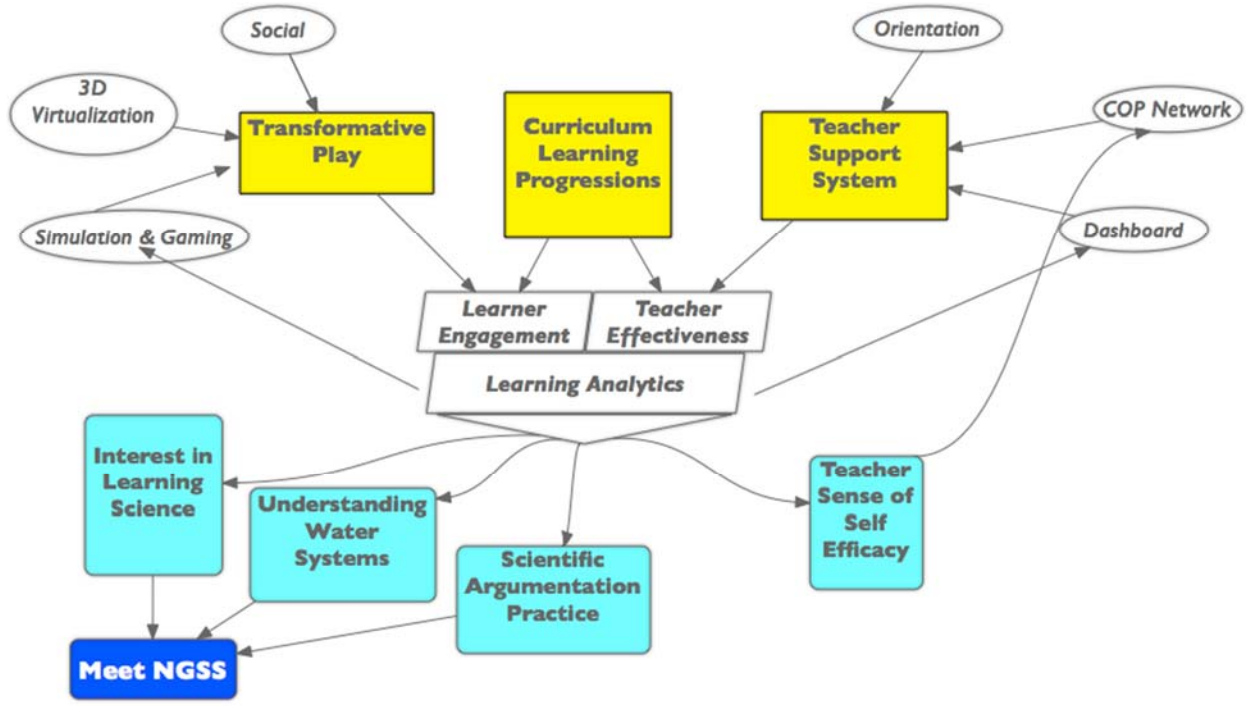
Contribution to Theory, Knowledge and Practice. MHS is based on two powerful conceptual frameworks of how to support learning: curriculum progressions and transformational play theory. For each of these frameworks MHS adds a component of adaptive and dynamic feedback to the student and teacher via learning analytics. Integrating predictive analytics into the teaching-learning process has great potential for enabling more personalized and appropriate learning for each student. Further, using gaming, virtualization and a teacher support system provides this personalization in a way that strongly engages students and can be implemented in distance learning. Additionally our model of game-based learning through simulation, transformational play and analytics can be disseminated to other design and development teams. We will use a popular game engine system and any new programming elements we produce will be made available via open source licensing. Finally since MHS is a software product it has the potential to be accessible to large numbers of schools in a relatively short period of time.

B. QUALITY OF DESIGN:

Goals. This project aims to achieve five goals in the service of developing and testing a scalable model for innovative and effective distance and blended virtual learning to meet the needs of learners in small and rural schools with diverse needs: 1) Develop a game-based 3D VLE for learning hydrologic systems and scientific argumentation. 2) Develop a learning analytics system to provide in-game assessment and feedback to students and enable teacher monitoring and

efficacy to intervene as needed. 3) Provide teacher professional development and support to ensure effective implementation. 4) Deliver and evaluate MHS in partner schools. And, 5) build new knowledge about game-based learning, analytics and teacher support for effective VLE.

Logic Model. The Logic Model below describes project inputs and activities leading to enactment of our theory of how learning is enabled in game-based 3D VLE and presents short-term, intermediate, and broad outcomes (more detail is available in Appendix. D).

Assumptions: Game-based 3D VLE engage and support students in developing and applying science competencies that meet the requirements of distance and blended learning. Learning Analytics supports teacher efficacy and appropriate intervention.
Inputs: i3 and matching funds, project staff expertise, partnerships, advisory board, advancing capabilities of technology, prior research on transformational play and curriculum progressions.
Activities: Iterative design and development process as represented in the 4 cycles
Outputs: Game-based 3D VLE experience for students and teachers implemented in schools meeting requirements for distance and blended learning. The key components of the system are: water science & argumentation simulation and learning systems using transformative play and learning progressions based curriculum, learning analytics system, and a teacher support system.
Enactment of Learning Theory and Short-term Outcomes: 

Intermediate-term Outcomes: Students meet NGSS and can make sense of big ideas of science, apply principles that cut across science disciplines such as causal reasoning and scale, and engage in sophisticated forms of scientific argumentation.
Long- term Outcomes: MHS serves as a scalable model for innovative VLE for distance and blended learning helping students in small and rural schools achieve to rigorous NGSS.

Activities. We plan 4 cycles of "design & development", "implementation & data collection", and "analysis & results." The first 3 cycles are focused on advancing and informing the design and development process and learning about how components of the system should work. The 4th cycle (to be described in the project evaluation section) quantifies how able teachers are to implement MHS and the impact on student engagement and understanding of constructs and competencies in hydrology and Scientific Argumentation.

Table 1. Four cycles of the iterative design and development process.

Cycle	Activity
1	**Usability** testing in years 1 & 2 to determine if students can use MHS as intended. Testing with 24 students at 6 time periods in university labs.
2	**Usage** testing in 1st half of year 3 to explore MHS use in school classes.
3	**Feasibility** testing in 2nd half of year 3 in distance or blended classes.
4	**Pilot Testing** in year 4 to evaluate student outcomes using pre and post indicators.

Sample for cycles 1-3: MHS is designed for middle and junior high schools students enrolled in earth and general science courses. Cycles 1-2 will recruit students and teachers in appropriate science courses in our LEA (Columbia Public Schools). These cycles focus on usability and human computer interaction. Meeting the requirements of these cycles is supported by our ability to observe and interact with participants during the learning. Cycle 3 will use 2 distance/blended courses recruited through the Blended Schools Network.

Cycle 1: In the first two months of the project the teams will undertake a design conference with

key project members and advisors to produce a Requirements Document for design and development. We will utilize a design conference format following innovative collaboration methods developed by BSCS (BSCS is a recognized leader in science education) to develop a framework and initial specifications for the curriculum, the software system and needed teacher support. Based on the requirements document the design team will develop scripts, prototypes and assessments for levels 1–4 which will then be reviewed by the Advisory Board (AB). Following any modifications the development team will produce levels 1-4. Usability testing with 12 students will take place in the Fall, 2015. The process of usability testing includes a trial of the levels at the Information Experience Lab (IE Lab) at MU, which is equipped to capture extensive data on the user experience. Usability tests in each cycle include having 3 students use the system, a review of findings, needed modifications, and a repeat of the usability test for changes needing retesting (which is expected through 4 turns of usability testing for levels 1-4 using different student testers for each turn). Following completion of usability testing of levels 1-4 we will follow the design process of creating scripts, prototypes and assessments for levels 5-8 as well as any modifications needed for levels 1-4. After the AB review, levels 1-8 will be developed and usability testing will be undertaken in the IE Lab in the Fall 2016.

For each usability turn in Cycle 1 the students will follow a think-aloud protocol with screen capture and audio-video recordings and be debriefed at the end of the session. The MHS software will also log student interactions and choices. We will apply qualitative methods to aggregate and synthesize feedback from the students to make judgments about HCI usability and engagement as well as the quality and sufficiency of missions, argumentation and gaming. Initial behavioral summaries will be organized as profiles for each participant and level. We will examine profiles for usability/engagement challenges that will then be prioritized using card-

sorting exercises. Case reports of argumentation will also be made for each participant and each level. The design team then reviews the results for each level and creates specifications for revisions before Cycle 2 implementation.

Cycle 2: Following Cycle 1 we will make whatever changes are needed to prepare all 8 levels for a field site usage test. The field usage testing will take place in the Fall, 2016 at a local blended science classroom over a 3 to 4 week period and assess not only student engagement in MHS but also test the supports available for the classroom teacher. The teacher supports will include the first edition of orientation and training materials and a dashboard for monitoring and intervening with students. At the conclusion of cycle 2 we will have determined, or need to make changes to assure, MHS is sufficiently engaging for sustained effort by students, all functionality is usable by the students and teachers, and data are appropriately captured and used for the analytics.

Cycle 2 tests how teachers and students use MHS in the context of a science class. Prior to implementation the AB, via a cognitive walkthrough, will be asked to rate usability and the tangibility of experience with science phenomena and practices. Short questionnaires administered after each online session will ask students to rate sense of presence and enjoyment with the learning experience. At the conclusion of cycle 2, a debriefing will be held with the teacher. Similarly, students will be invited to participate in a discussion in order to identify strengths and limitations of the current implementation. Profiles for each participant and level will be written and usage/engagement challenges will be analyzed with card-sorting techniques.

Cycle 3 explores the feasibility of MHS in the online context of distance and blended courses. The two teachers from BSN schools who are participating on the design team will implement MHS with their online students. For cycle 3 our teacher team and the BSCS science educators will use data from cycle 2 to complete the design of the teacher orientation and training materials

as well as the teacher dashboard for representing learning analytics. They will also specify a networked environment to support teachers during implementation. The Network will enable discussion and sharing among teachers and an MHS moderator. The AB will review the dashboard, orientation and network to identify needed improvements prior to the feasibility test.

Based on qualitative findings from cycles 1 & 2 we will develop a “game play and progress” coding scheme for characterizing each action and opportunity for action in MHS (e.g., response to NPC question, assertion or direction; interaction with data cards; movement in world; accessing learning materials, etc.). Our research team has completed analogous work for the iSocial system (Schmidt, Laffey et al, 2012). These codes will be extended as we examine learner activity in cycle 3 creating a coding scheme for characterizing student play and progress. A close comparison of the behavioral traces gathered by the MHS system and the results of the qualitative analysis will be made, with the aims of a) proposing a set of data to be automatically gathered and analyzed, and b) identifying a set of indicators for engagement, progress in understanding hydrologic systems, progress over the 10 levels of argumentation (Osborne et al, 2013) and game play. This process will iteratively improve how log data can be used to analyze progression (or lack of it) within and through levels. Highly accurate learning analytics are not likely to be achieved until much larger samples of students engage in MHS but these efforts should enable a framework to be developed and refined (for more discussion see the methodological description for developing Learning Analytics in section 1 of Appendix J). In addition qualitative methods to aggregate and synthesize reviewer responses, student responses, discussion contributions and the teacher debrief will be applied to create a baseline regarding the quality of engagement, functionality, system performance, and usability. The analysis identifies emerging issues about the feasibility of implementation. These results are specifications for

revisions before pilot test implementation.

Output. *Outputs include learning systems, analytics, and teacher support systems.*

Water science & argumentation simulation and learning systems: The learner experience of these systems has been briefly characterized in the Significance section. At the completion of the design cycles we will have produced approximately 16 hours (4 weeks) of online instruction in an 8-level game. The first 6 levels are intended as individual play where the student works through game challenges while learning hydrologic systems content and building argumentation skills and competencies. As students progress they see their scores rise, gain badges and powers in the game, as well as receive affirmation from the NPCs in the game and their teacher. Students who struggle will repeat the game play and may receive interventions outside the game from their teachers or inside the game based on analytics or by in-game interventions the teacher can enact such as sending a support NPC to help the student. Table 1 illustrates the hydrsocience and argumentation objectives at each game level.

Table 2. MHS levels with proposed tasks and argumentation competencies.

Level	HydroScience Task/Problem	Argumentation (Key features)
1	Create a watershed	Supporting claims with evidence
2	Track a pollutant through the watershed	Supporting claims with evidence Linking claims and evidence with reasoning
3	Explore the impacts of vegetation & substrate porosity	Constructing a complete argument
4	Place a well	Providing an alternative counter argument
5	Alter precipitation rates	Providing a counter-critique
6	Establish a human community	Constructing a one-sided comparative argument
7	Create a system for treating polluted/contaminated water	Providing a two-sided comparative argument
8	Water wars between communities	Constructing a counter claim with justification

Learning analytics system: The principal goal of the analytics system in MHS is to help instructors and students notice when behaviors in the system lead to learning, or represent critical decisions or wayfaring markers on the path to learning. Identifying behaviors that correspond with learning in a VLE is a complex endeavor that cannot be disembodied from the curriculum. Goggins' work developing a methodology and ontology for making sense of behavioral traces across a range of work and learning environments, called Group Informatics, provides a systematic foundation around which we will build our design iterations for MHS Analytics (Goggins, Mascaro and Valetto, 2013). The first step in developing any behavioral analytics system is abstracting the levels of organization within the system – the part of the trace data from which context can be reconstructed (Goggins & Petakovic, 2014). What traces do we have to work with by default? What other behavioral signals do we need to capture? And, most critically, how are a user's specific actions in the VLE connected to the curriculum being followed; what signals success? Frustration? Loss of interest? As the analytics lead researcher for Virtual Math Teams with Geogebra (VMTwG), Goggins developed a systematic approach for applying activity theory as a lens for processing raw traces, directly connected to curricular aims, showing statistically significant connections between group performance and participation on specific curriculum topics in VMTwG (Xing, Wadholm & Goggins, 2014). Further developing this approach, Xing, Petakovic & Goggins (2014) employed a genetic programming algorithm against activity theory so as to inform and reshape trace data to predict individual performance in a more user-understandable visualization in the VMTwG environment.

For MHS analytics we will follow a similar, systematic path as was used in VMTwG. We will develop analytics that are focused on tracking individuals' specific choices and analyzing those discrete choices against a backdrop of learning progressions in water systems and

argumentation. Along with tracking choices, our approach will track and systematically analyze inscriptions (Borge & Goggins, 2014) created and used by students within the environment.

This process of making learning visible is most helpful for student success if it shapes how teachers use their time and energy in support of students. For example, the planned analytics can provide a teacher, via the dashboard, the capability to provide scaffolding for a student exhibiting problems by inserting a supportive NPC tailored to the student struggles into the student's game play to help avoid failure and frustration when the student returns to MHS.

Teacher support system: In both face-to-face and online environments, teachers shape how curriculum materials are enacted (Ball & Feiman-Nemser, 1988; Forbes & Davis, 2010). If a teacher's understandings and beliefs about instruction align with the philosophy of the curricula, then it is likely that there will be a synergistic relationship between use of the materials and practice. In MHS many curricular decisions (that might be made by teachers in other kinds of learning environments) are determined by designers of the VLE and by analytics, but teachers play a critical role in interpreting student progress, providing feedback, and ensuring learning. Therefore, MHS teachers will need to understand the MHS teaching and learning model, the analytics system, and ranges of responsive actions and feedback that can be offered to learners engaged in the environment. To support coherent implementation and curriculum appropriate responses and actions by the teacher, we plan to develop 3 forms of teacher support through the design cycles to be pilot tested. The first form are training materials and simulations that prepare the teacher to understand the hydrology and argumentation content of the game, the game play that students will experience and the processes and tools available to the teacher during student game play. To make the materials educative, we will integrate a variety of teacher supports that align with the heuristics suggested by Davis and Krajcik (2005). The second form is a networked

community of practice consisting of other teachers using MHS or artifacts from teachers who have used MHS and moderators from the MHS team. The third form of support for effective teaching is a dashboard representing data about the students during gameplay. This dashboard provides the teacher an up-to-date report from the learning analytics system as well as a recommender system for when and how to intervene for students. The dashboard will also provide ”Just in Time” introductions to lesson activity and on-demand resources such as FAQs.

Potential Risks: Of the many risks in the development of a complex learning software system, 3 are most notable. First, there are many pedagogical, artistic and technical choices that must be executed well to develop a powerful learning system. The primary approach to addressing this challenge is the iterative process and benefit of an advisory board to help utilize the systematic feedback designed into our project. The formative evaluation from our independent evaluator should help assure a comprehensive view of how well we are executing toward meeting objectives. A second risk is that if the Learning Analytics are not effective, much of the pedagogical power of MHS and support for the teacher will be diminished. There are many challenges such as which data to mine, which processes to apply to the data and how to represent the results in meaningful ways to students and teachers. However, developing Learning Analytics is an iterative process where more trials can provide better results. Our plans are to continue data collection and analytics refinement after the i3 project is ended and MHS starts to be used in schools. The third risk is that teachers may feel overwhelmed by the new role and teaching tools/resources and fail to understand and carryout some aspects of the teaching. This is partially mitigated by designing the student experience to allow most students to succeed even if there is little teacher intervention. We will also use the analytics to examine teacher behavior in MHS so as to provide feedback to the teacher as well as notify the MHS network moderator if

proactive support is needed in some cases. In addition BSCS and BSN both have strong backgrounds in supporting teachers and will help drive these efforts as well as identify corrective approaches to the extent we see problems.

C. MANAGEMENT PLAN & PERSONNEL: The project milestones chart shows the key milestones in each cycle (C) and leadership responsibilities (Resp) for faculty and staff of MU (MU), staff from BSCS (BSCS), Advisory Board (AB), staff of Blended Schools Network (BSN), Missouri Partnership for Educational Renewal (MP), Columbia Public Schools (CP), Faculty of Wright State (WS), and the project independent evaluator, Thomas Reeves, (TR).

Project Milestones	**C**	**Resp**	**2015**			**2016**			**2017**			**2018**		
			Sp	**Su**	**Fa**	**Sp**	**Su**	**Fa**	**Sp**	**Su**	**Fa**	**Sp**	**Su**	**Fa**
Design conference – Requirements Doc + Annual Formative Review	1	MU	*			*			*			*		
Develop & validate instruments for assessing understandings of water systems and argumentation	1 2	WS	*	*	*	*	*	*						
Design Curriculum for MHS levels 1-4	1	BSCS	*	*										
Design and develop MHS levels 1-4	1	MU	*	*	*									
Advisory Board reviews levels 1-4	1	AB		*										
Recruit students for Usability testing	1	CP		*										
Usability testing of MHS levels 1-4	1	MU			*									
Design Curriculum for MHS levels 1-8	1	BSCS			*	*								
Design and develop MHS levels 1-8	1	MU				*	*	*						
Advisory Board reviews levels 1-8	1	AB					*							
Usability testing of MHS levels 1-8	1	MU						*						
Examine data and make revisions to MHS	2	MU						*	*					
Develop Teacher Support System (TSS)	2	MU BSCS				*	*	*	*					
Advisory Board reviews MHS-TSS	2	AB							*					
Usage testing of MHS	2	MU							*					
Examine data and make revisions to MHS	3	MU							*	*				
Advisory Board reviews MHS-TSS	3	AB								*				
Revise Teacher Support System (TSS)	3	MU BSCS							*	*	*			
Feasibility testing of MHS	3	MU BSN									*			
Develop and Audit comparison curriculum (standard units from BSN) and	4	BSCS								*	*			

Project Milestones	C	Resp	2015			2016			2017			2018		
modify to align with MHS														
Recruit students for Pilot study	4	BSN MPER									*			
Examine data and make revisions to MHS	4	MU							*	*				
Advisory Board reviews MHS-TSS	4	AB								*				
Revise Teacher Support System (TSS) including updated Learning Analytics	4	MU BSCS									*	*		
Revise all levels of MHS	4	MU									*	*		
Pilot study	4	MU BSN										*	*	*
Analyze pilot study data and prepare publications	4	MU WS											*	*
Disseminate findings through presentations & publications	4	MU			*			*			*			*
Formative Review by AB and Evaluator	4	AB				*			*			*		
Summative Evaluation	4	TR WS										*	*	*

James Laffey of the MU faculty and Mark Bloom from BSCS will provide overall management and oversight for all activities and collaborative efforts. Please see letters of commitment in Appendix G from advisory board members, project partners, independent evaluator, and our partners BSCS, BSN, WS, CP, and MP. The letters from BSN and MPER, two large organizations that are stakeholders in supporting rural schools, indicate the need for systems like MHS and their commitment to the project. Within the first 2 months of each year the team will have a multi-day conference. In the first year the conference will focus on building the team, establishing networked communication protocols and producing a requirements document for design and development. In years 2-4 the conference will be used for review and improvement to the requirements document and formative evaluation of progress towards performance targets. The principle methods and metrics for annual review of project performance are: (1) The project team will complete a cognitive walkthrough of the state of MHS and review data from the testing processes of the previous year's cycle, (2) In Yr 2 & 3 profiles of participant performance and usability issues for levels will be used to judge the adequacy of usability/engagement and how

well MHS meets expectations set out in the requirements document. (3) In year 4 the team will undertake a cognitive walkthrough of MHS and the teacher support system while reviewing data from the usage and feasibility studies. And (4), The team will use data from the usage and feasibility studies to judge the adequacy of data being used for Learning Analytics and whether feasibility has been established. In addition, Tom Reeves, the independent evaluator will attend each year's conference and provide a formative report about progress toward project goals.

In addition to the MHS team's annual review, the project also has an advisory board that will provide external feedback for consideration in the annual review and as a direct contribution to the design process in each cycle. The Advisory Board (AB) includes researchers and educators with diverse expertise (see personnel) and provides review on five occasions across the 4 cycles. At each cycle the AB will follow a structured process to provide feedback appropriate to the objectives of the cycle. For example in the usability cycle the AB will be asked to rate usability/engagement and the tangibility and adequacy of experience with science phenomena.

The annual conferences, the formative evaluation report from the evaluator and the input of the advisory board will provide systematic review and feedback for continuous improvement. Project success also requires communication systems for regular discussion and efforts to meet emergent challenges. The project will use email and twitter for daily interaction, use Dropbox for file sharing, a website for discussion boards and sharing news outside the project team and Zoom for Internet conferencing. Within functional teams for design, development, assessment, teacher support, analytics and testing we will have biweekly meetings with reports posted to discussion boards on the website and Dropbox.

PERSONNEL: The MHS project team consists of a strong lineup of researchers, schools, and technology partners with expertise in game design, STEM education, learning analytics, teacher

professional development, and evaluation. Laffey, Sadler & Bloom have all led large, multi-year federally funded projects. Profiles for key personnel are provided below and biosketches are provided in Appendix F.

James Laffey (PI) is a Professor in the School of Information Science and Learning Technologies at MU. Prior to MU, Laffey worked for Apple Computer, Inc. conducting research on learning and support systems and developing award winning interactive learning systems. Since coming to MU, Laffey has been the PI for several large Department of Ed and NSF grants totaling nearly $8.0 million including a just completed Goal 2 award from IES for *iSocial,* a VLE for youth with Autism Spectrum Disorders. *Laffey will provide overall project management and lead the development and testing of the 3D virtual learning environment.*

Troy Sadler (Co-PI) is a Professor of Science Education at the University of Missouri (MU) and Director of the *ReSTEM Institute: Reimagining & Researching STEM Education.* He recently directed a NSF-funded project to design and study *Mission Biotech* (MBt), a virtual environment for science learning in the context of biotechnology. Sadler's research focuses on engaging scientific argumentation and *he will lead curriculum design and assessment efforts*.

Sean Goggins (Co-PI), Assistant Professor in the MU iSchool, is a leader in the NSF sponsored research coordination network for Digital Societies, and the PI for 3 NSF grants focused on the design and development of analytics systems for technology mediated work and learning. *Goggins will lead learning analytics development*.

Mark Bloom (Co-PI) is a Science Educator at BSCS where he has directed development of 12 print- and web-based curriculum modules for middle and high school students. Bloom has experience organizing advisory and design conferences, implementing nation-wide field tests, and leading professional development. *He will lead the teacher support system development.*

William Romine (Co-PI) serves as an Assistant Professor within the Department of Biological Sciences at Wright State University undertaking research related to educational measurement and assessment within innovative instructional contexts. *Romine will lead assessment development efforts and analyses for the pilot study.*

Thomas Reeves (Independent Evaluator) is a Professor at the University of Georgia and an expert in evaluation of technology-based learning systems and design-based research. Reeves has conducted extensive evaluations for organizations such as CDC, U.S. Army, and the World Health Organization, IBM, Apple, and AT&T..

The Advisory Board members were selected based upon specific areas of expertise related to the project. **Beth Covitt** (U. of Montana) specializes in teaching, learning and assessment of water systems science. She is one of the lead researchers of the water systems learning progression central to the MHS design. **Douglas Clark** (Vanderbilt) is a learning scientist and science educator who specializes in creating and studying technology-based environments for supporting science learning. **Victor Sampson** (U. of Texas) is a leading researcher in the area of scientific argumentation. We plan to leverage lessons learned by Sampson and his team in the IES-funded *Argument Driven Inquiry* project, which serves as the basis for our project's NGSL assessment. **Krista Gaylen** is a curriculum designer for online learning with the MK12, the MU online school and has extensive experience developing and testing VLEs. In addition to these formal advisors, our team works closely with **Jed Friedrichsen**, CEO of the Blended Schools Network, **Michael Szydlowski**, the Science Coordinator for Columbia Public Schools, and **Dan Lowry**, co-director of the Missouri Partnership for Educational Renewal (MPER)

Table 3. Project Responsibilities of Additional Key Roles

Contributor (affiliation)	Project Responsibilities
Ryan Babiuch (MU)	Lead programmer for the MHS environment.

Betty Stennet (BSCS)	Support curriculum development and creation of teacher support materials.
Christopher Wilson (BSCS)	Support development of assessments; Consult on research design
Susan Kowalski (BSCS)	Lead development of teacher support materials and consult on research design.
Game Designer	TBH, likely hired from graduating PhDs in Comp Sci or Learn Tech
Modeling Specialist	TBH, likely hired from graduating PhDs in Comp Sci, Learn Tech or Arts

E. PROJECT EVALUATION: Project Evaluation will include formative and summative components. Formative evaluation is included in the design plan with iterative design, review, and testing as described for cycles 1-3. The advisory board provides a critical examination of products and processes to insure best practices and continuous progress. The annual review with a formative report on progress toward objectives by the external evaluator also assures attention to performance targets. Summative evaluation will be undertaken with an impact study consistent with research type 4 of the IES/NSF Common Guidelines. We plan for the student learning experience to represent a "typical" online experience but the project team will help schools assure sufficient technology infrastructure and provide teacher training and support. For the pilot the teacher support system will be provided by BSN with backup from MU and BSCS.

Impact Study Research Questions:

Do middle school and junior high school students learning with MHS:

(1) demonstrate greater water systems knowledge, scientific argumentation, next generation science learning, and interest in science, than similar students in a comparison condition?

(2) gain significant water systems knowledge, scientific argumentation, next generation science learning, and interest in science, after receiving MHS?

Methods: We will use a cluster random assignment design for the impact evaluation. Recruitment and selection will be conducted in partnership with the Blended Schools Network (BSN) and the MU Partnership for Educational Renewal (MPER). BSN provides distance and

blended learning solutions for small and rural schools across the country, with 77,000 students; 11,000 teachers; and 169 schools in nine states. MPER is an organization with 22 member school districts in Missouri representing over 180,000 students. While the specific schools that will participate in the pilot are not yet identified all participating schools will be in districts listed as rural or town-remote by NCES in Fall 2017. BSN and MPER will identify eligible districts and send invitation letters to district science coordinators soliciting their interest and participation. Once districts have committed to the project and qualified a school the school will be randomly assigned to a treatment group. Specifically we plan to recruit middle school science classes for the following: 20 classes of DL students expecting 10 students per class, and 40 classes of BL students expecting 25 students per class. In order to serve more MHS students, we use an unbalanced design as shown in the table. The total sample includes 840 MHS students and 360 comparison students.

Table 4. Unbalanced student distribution by treatment/comparison and distance/blended

	MHS Treatment	Comparison Intervention
Distance Learning Classes	140 students (14 classes)	60 students (6 classes)
Blended Learning Classes	700 students (28 classes)	300 students (12 classes)

Classes assigned to the comparison group will be offered an opportunity to implement MHS in the Fall of 2018. Teachers will be trained and schools prepared for their treatment option. Pre measures will be taken and the 4-week treatment period will be undertaken between February 1 and April 30. Following completion of the treatment post measures will be taken for student outcomes and for teacher efficacy. The comparison classes will undertake a common 4-week curriculum based on an existing water science curriculum developed by BSN. We will adapt the BSN water systems curriculum to insure that it meets all of the water system learning objectives

to be assessed on the pre and post-tests. There is no alternate curriculum for the scientific argumentation objectives of MHS but post testing will show the extent to which practice using water systems information in MHS yields different levels of argumentation competency. Time on task measures will be taken for both MHS and the comparison to understand the effect of effort and time on outcomes.

Power analysis: The above sampling plan includes 60 classes (average class size of 20) with 70% of classes assigned to MHS. Assuming the intra-class correlation (ICC) is 0.20, the proportions of variance explained by the pretest and demographic information are 50% at level 1 (students) and level 2 (class), this sample size for the two-level cluster random assignment design will have a minimum detectable effect size (MDES) of 0.28 for a two-tailed test with an alpha of 0.05 and statistical power of 80% (Dong and Maynard, 2013). An effect size of 0.28 is equivalent to moving a student at the 50th percentile up to 61th percentile.

Measures: (for a more complete description of the measures see Appendix J-section 2)
The *Independent Television Commission's-Sense of Presence Instrument* (ITC-SOPI, Lessiter et al, 2001) will be used to assess sense of presence in MHS. The ITC-SOPI is a validated self-report questionnaire that can be used to measure four factors: 1) Spatial Presence 2) Engagement 3) Ecological Validity; and 4) Negative Effects.

The *Student Interest in Technology & Science* (SITS) instrument (Romine, Sadler et al., 2014) will be used to assess interest. The SITS is made up of four related sub-constructs: interest in learning science, interest in learning science with technology, interest in careers in science, and interest in careers in technology.

Water Systems and Argumentation instrumentation will be developed as part of the project work. For measuring participants' understandings of water systems, we will use assessment tasks that

have been developed in the context of the water systems learning progression cited in Table 1 (Gunckel et al. 2012). In years 1 & 2 we will pilot test the items and associated rubrics with student samples from the target population to ensure that the items collectively yield scores of sufficient validity and reliability for understanding the impact of MHS on student understanding of hydrologic processes. We will take a similar approach for assessing argumentation competencies. The research group responsible for the argumentation learning progression has developed tasks consistent with the empirically validated progression (Osborne et al., 2013; 2014). We will use these tasks as the basis for our argumentation instrument and will make modifications based on pilot testing during years 1 & 2 of the project.

The assessment of NGSL requires an instrument that challenges learners to negotiate water systems science in the context of argumentation. The format for this assessment of NGSL and the associated scoring rubrics are based on assessment innovations made in the IES funded *Argument Driven Inquiry* program (Walker & Sampson 2013; Sampson et al., in review).

Teacher self-efficacy; As part of post testing we will use two scales modified for teaching with MHS from the Teacher Sense of Efficacy Scale- Short Form (TSES; Tschannen-Moran & Woolfolk Hoy, 2001), to assess efficacy for student engagement and for instructional practices.

Analysis: For impact analysis research question 1, we use two-level Hierarchical Linear Model (Raudenbush & Bryk, 2002), where students are nested within classes. We elaborate the models below.

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 (class): $\begin{array}{l} \alpha_{0j} = \beta_{00} + \beta_{01}(MHS)_j + \beta_{02}(DL)_j + u_{0j} \\ \alpha_{mj} = \beta_{m0}, m = 1,\ldots,M \end{array},\ u_{0j} \sim N(0, \tau^2)$

Where y_{ij} is the outcome variable for student *i* in class *j*; X_{mij} represents *M* student-level covariates including pretest and demographic information; $(MHS)_j$ is a binary variable indicating treatment condition (*MHS* = 0 for non- *MHS* class; *MHS* = 1 for *MHS* class); $(DL)_j$ represents the types (*DL* = 0 for BL class; *DL* = 1 for DL class). The effects (β_{m0}) of student-level covariates are assumed constant across classes. β_{01} represents the average effect of *MHS*. For impact analysis research question 2, the above two-level Hierarchical Linear Model can be modified to examine if the gain score is statistically different from 0 for the treatment sample, where y_{ij} now represents the gain score between posttest and pretest of the outcome variables for student *i* in class *j*; X_{mij} represents *M* student-level covariates including demographic information; no $(MHS)_j$ or $(DL)_j$ will be included in level-2 equations. β_{00} estimates the average gain score.

Measurable Threshold for Implementation: The minimum effect size estimates are calculated with PowerUp! and reflect the following basic expectations: power (□) = .80, significance (□) =.05, and intra-cluster correlation coefficient (icc) = .20. The first two are standard, and the third is empirically-based on research in education (e.g., see Hedges & Hedberg, 2007). Estimates vary based on the projected variation explained (r^2) by the model and whether models are estimated with fixed or random effects. As per the NEi3 analysis and reporting guidelines, annual thresholds (see below) are set for each key component as depicted in the logic model. If the fidelity measure is greater than or equal to the threshold, the key component will have been implemented with fidelity; if it is less than this threshold, the key component will have not been implemented with fidelity.

These thresholds are outlined as follows:

(1) For the MHS treatment group: the student completes at least 80% of the MHS as measured by advancement of levels documented within the game data logging system.

(2) For the control group: the student completes at least 80% of the comparison curriculum as measured by completed exercises.

(3) For teachers in both the treatment and comparison groups at least 80% of the teachers will rate efficacy on the TSES above the mid point of "some influence" for items

Resources for Carrying out the Evaluation: BSN and MPER provide access and credibility for recruiting rural school districts. BSN manages a large network of Internet services and resources and is well equipped to provide access to MHS, connections for online data collection and technical support to the participating schools. The MU and BSCS staff will provide backup to the BSN team and provide the content and support for teachers. A SWAT team from BSN, MU and BSCS will address any school or teacher that requires troubleshooting to startup or sustain the pilot implementation. While the data from analytics are being generated and collected Goggins will provide quality assurance. Sadler will manage the implementation and collection of pre and post assessment measures and Romine will organize and provide descriptive statistics. Each of the first 3 years the Independent Evaluator provides a formative evaluation. In the pilot test year he will have 3 primary tasks: (1) a visit plan for 10% of the sites that will include interviewing teachers, administrators and some students and parents as well as observing operations, (2) a debrief plan for all sites, and (3) directions for analysis of the instrumentation and analytics data to address all evaluation goals. Professor Reeves' contract for the year of the pilot study assures approximately 40% of effort and his work is supplemented by Romine for undertaking statistical analysis and by 2 graduate research assistants for evaluation tasks and qualitative analysis.