TABLE OF CONTENTS

Scaling and Sustaining Mission HydroSci: Game-Based Learning for Next Generation Science Learning

A. SIGNIFICANCE – Absolute Priority 1: Strong Evidence; Absolute Priority 3: STEM

Mission HydroSci (MHS) (Laffey et al., 2019; Laffey et al., 2016) is a 3D game-based learning (GBL) environment developed with I3 grant support for teaching key science concepts and scientific argumentation skills. MHS is aligned with the Next Generation Science Standards (NGSS, 2013). Students engage in scientific argumentation and integrate practices of argumentation with disciplinary core ideas and crosscutting themes in middle school science. Running on high performance Macintosh and Windows (Mac/Win) systems, MHS replaces 6 to 8 class sessions in science courses that address general and earth science. To achieve these student learning objectives, MHS incorporates several system elements: gaming, 3D virtualization, a problem-solving context, social interactions, empirical learning progressions, scaffolding for argumentation, an empirically-grounded instructional model, and an analytics system (more about MHS can be found at MHS.missouri.edu and in Appendix J.1).

Contributions to new knowledge and understanding effective strategies: Half (164 million) of Americans play video games and 74% of parents believe games can be educational for their children (ESA, 2019). Along with most educators and the general public, we see children comparing the engaging, active experience of playing games with the often placid and passive nature of being in a classroom (Qian & Clark, 2016; Plass et al., 2020).

MHS meets important NGSS outcomes desired in schools, and while games hold great promise for supporting students to important learning outcomes and many teachers are starting to use games in their classrooms, most games in schools are short-form games used to simply introduce a topic or provide drill and practice (Takeuchi & Vaala, 2014). Scaling and further evaluating MHS will build new knowledge and advance important strategies for teaching with technology and game-based learning (GBL). Effective use of educational technology has long been an important goal of national policy (NETP, 2017) but is even more urgent in the wake of the COVID-19 lockdown and defining what the new normal will be for teaching. Certainly, there is already a greater emphasis on online instruction. This project will also build understanding about how to engage students, especially those who may be uninterested or high need, in science education and will build approaches for using analytics to support teaching.

B. QUALITY OF PROJECT DESIGN

B1. Goals, Objectives, Outcomes for Enhancing and Scaling MHS: The proposed five-year project has two goals: (1) improve the scalability and sustainability of MHS by improving components of MHS and expanding the platforms for accessing MHS to include lower performance Mac/Win computers and iPads; and (2) evaluate, extend, and disseminate MHS and the MHS evidence-base for (a) improved student academic achievement for middle school science education, (b) better understanding of engagement of students who are at-risk in science education, and (c) cost effectiveness and teacher efficiency in using MHS.

Exhibit 1: Goals, Objectives, Milestones, Outcomes, and Measures for Scaling MHS

Goals & Milestones (MS)	Outcomes	Measures
Goal A: Improve the scalability and sustainability of MHS		
Obj 1: Increase the number of schools that can use MHS by expanding the platforms that can be used to implement MHS.		
MS 1.1 Optimize the MHS Mac/Win applications to run on lower performance computers	Proof of Concept (POC) with 30 Field Test (FT) teachers successfully using optimized MHS Mac/Win version with classes	FT Teachers meet threshold for use of optimized Mac/Win version (80% of students complete all units of MHS)
MS 1.2 Port a new version of MHS to run on iPads	POC with 30 FT teachers successfully using iPad version with classes	FT Teachers meet threshold for use of iPad version (80% of students complete all units of MHS)
Obj 2: Plan for scale up by improving efficiency in accessing MHS and by disseminating information for further development, usage, and replication		
MS 2.1 Develop myMHS, a comprehensive teacher and student registration component of MHS for use in Pilot Study	- 8 teachers implement pilot versions of myMHS and MHS - Deficiencies identified	Students registered and meet threshold for use

MS 2.2 Implement myMHS for self-registration, access to MHS and support during field test	POC 60 FT teachers successfully use myMHS to register students	All students registered with < 2% errors. 90% high rating on teacher form.
MS 2.3 Disseminate findings, implications, and MHS	Reporting about and access to MHS for educators, researchers, and developers	Annual report on dissemination indicates 100% completion of plan activities per year
Obj 3: Improve the effectiveness and efficiency of using MHS for teachers by enhancing the Teacher Support System (TSS)		
MS 3.1 Develop TSS: Orientation course + guide + community + coaching for use in Pilot Study	- Materials developed and reviewed - 8 teachers complete TSS in pilot - Deficiencies identified	Counts (log data of TSS) of teacher participation in TSS including community
MS 3.2 Implement TSS for teacher support during field test	POC 60 FT teachers complete TSS	Surveys of teachers report high teacher satisfaction with materials in guiding use of TSS
Obj 4: Continuous improvement of MHS throughout the project		
MS 4.1 Core group of local teachers (CIC classes) implementing MHS	Interviews w/ teachers & focus groups w/students	Annual report - Improved efficiency of using MHS
MS 4.2 Learning Analytics (LA) development work	- Interviews w/pilot teachers -Analytics data/outcomes	Annual report + Dashboard updates
MS 4.3 Embedded assessment of engagement	- Interviews w/ teachers - Embedded assessment data/outcomes	Annual report + embedded assessment updates
MS 4.4 Plan and capacity for sustainability	Starting Year 1, work with MU Technology Transfer program to articulate a plan for sustainability	- Reviews in Year 1, 3, and 5 by experts/ sustainability - MHS signups for 25-26 academic year (post grant)
Goal B: Evaluate and extend the MHS evidence base		
Obj 5: Impact Study of MHS with (60) teachers in Years 3, 4, & 5 of grant		
MS 5 Recruit, randomize, implement field test, and analyze	Data on outcome measures via randomized block design	Water Systems and Argumentation Assessment measures
Obj 6: Understand and explore implementation issues of MHS		
MS 6.1 Finalize measure of fidelity of implementation that aligns with logic model and meets EIR criteria	Pilot fidelity measurement using data from TSS and MHS logs during field test	Teachers meet threshold for high level of fidelity of use (80% of students reach unit 4)
MS 6.2 Assess teacher attitudes about desirability to use MHS in future	Post MHS field test interviews with teachers	Content analysis of interviews and report on findings of teachers rating of desirability of use of MHS
MS 6.3 Assess student engagement with MHS	- Focus groups of students in CIC classes - Student Questionnaire from field test - Embedded assessment from field test	- content analysis of focus groups - student ratings -algorithms from log data Report on findings of extent to which teachers have high rating of desirability of use of MHS
Obj 7: Understand, explore, and plan for scale-up of MHS		

MS 7.1 Establish impact of moving to tier 2 (reduced level of support) of TSS	Data on outcome measures + fidelity in comparison of Year 1 teaching with Year 2 teaching	- Water Systems & Argumentation Assessment measures - Tier 2 Teachers meet threshold for use (80% of students complete all units of MHS)
MS 7.2 Identify cost effectiveness of MHS	Collect data for IES cost analysis protocol	Cost Out Tool analysis

B2. Project is Appropriate and will Meet the Needs of the Target Population: Twelve of the 13 teachers from the field test for the development project said they would like to use MHS again. We saw this as extremely positive given that many teachers and their classes struggled with numerous technical glitches associated with low performing computers. While the teachers and their classes enjoyed MHS, the primary reason teachers wanted to continue using MHS was because they saw how it engaged their students and in particular engaged students who were not usually highly engaged in science. One teacher expressed her enthusiasm for using MHS as:

> *I would say that we had almost 100% engagement while playing the game almost every day that we did it. It was also interesting to watch some of the kids help other students…so there was just a lot of excitement that came with playing the game.*

Another teacher referenced how students who typically do not do well can become leaders.

> *The cool thing about the game was that some of the boys that maybe aren't so good at paperwork. They took off with this and were like the leaders. They got to help other kids when they were having problems. They became the leaders and that was neat.*

Lee et al. (2016) documented that, in middle school, students lose their interest in science and achievement drops. Bathgate and Schunn (2016) summarized a broad array of research:

> *Many studies note that science interest is especially sensitive during adolescence, as there is a frequent drop in science interest during this age (Bryan et al. 2011; Gottfried et al. 2001; Hawkey and Clay 1998; Osborne et al. 2003; Simpson and Oliver 1990). ….*

But it is also worth noting that particular experiences can lead to increases in interest (e.g., Hulleman and Harackiewicz 2009) and some environments can prevent declines in interest (Vedder-Weiss and Fortus 2011). (p. 424)

In addition to being engaging for students, MHS addresses an urgent need for curriculum material aligned to NGSS. In a survey of 710 science teachers (Haag & Megowan, 2015) from 38 states, middle school teachers reported the need for professional development and curricular materials to teach NGSS-aligned science and argumentation. Further, a needs assessment (Harri et al., 2017) with 214 schools from 16 states concluded that teachers need additional support for enacting NGSS; in particular, they need access to teaching resources well aligned with NGSS.

A successful implementation of MHS can lead to growth of GBL across science education and potentially broadly impact engagement, NGSS and general science knowledge. We have evidence that MHS can significantly strengthen the teachers' repertoire of strategies for engaging students and teaching water systems and scientific argumentation. Research coming out of the project will also improve understandings about the use of analytics in GBL and how to cost effectively support teachers using GBL.

B3. Proposed Activities Constitute a Coherent and Sustained Program R&D: Projects in addition to MHS, such as *iSocial* (Laffey, Stichter, & Galyen, 2013), *River City* (Clarke et al., 2006), *Mission Biotech* (Sadler et al., 2013), *Quest Atlantis* (Barab, Sadler et al., 2007), *EcoMUVE* (Metcalf et al., 2009), *Crystal Island* (Rowe et al., 20011) and *SimCityEDU* (Glasslab, 2013) have begun the process of taking on the challenge to demonstrate and build new knowledge about how 3D GBL engages students and can produce significant student learning outcomes. A review of the role of games and simulations in science education (NRC, 2011) suggests that these technologies may be an important approach for achieving NGSS-aligned learning. The report concludes that the evidence for games is still emerging. More recent reviews (Connolly et al., 2012; Boyle et al., 2016; Clark et al., 2016) report many studies showing digital games enhancing student learning relative to non-game conditions. However, a deeper analysis by Clark (2016) shows that almost all the studies focused on lower-order learning outcomes. *In addition, a review of the WWC shows no other GBL interventions in science education with significant positive effects.* The strong evidence shown by the evaluation of MHS suggests that it could be a model for other game-based online learning systems, including (1) assessing and meeting higher-order learning standards aligned with the NGSS, (2) implementing innovative approaches in the use of learning analytics for continuous improvement of game play, as well as helping the teacher to be an observer and supporter of student performance (Goggins et al., 2013; Laffey et al., 2011), and (3) exploring how best to support teachers when using learning games.

Our project builds from extensive research related to students' understandings of water systems (e.g., Dickerson & Dawkins, 2004; Endreny, 2010; Shepardson et al., 2007). Using a previously developed and empirically-based learning progression for water systems and associated assessments (Covitt et al., 2009; Gunckel et al., 2009: Sadler et al., 2017), we have integrated the learning progression for teaching water systems with an explicit focus on argumentation (as can be seen in the intro document in Appendix J.1). Argumentation has long been recognized as a central epistemic practice within science (Kelly & Takao, 2002) and as an important goal for science education (Newton et al., 1999). MHS implements scientific argumentation not as a simple skill, but rather as a competency that necessarily requires reasoning and application of content knowledge (Osborne et al., 2013). Learning Analytics (LA) from extensive logs of in-game student behavior are used for continuous improvement of MHS but most directly and practically to make the teacher a partner in the game play. Teachers have dashboards for each class that present to them each student's progress and some key performances such as how many tries a student took to complete an argumentation task. Co-PI Goggins is using MHS to develop new machine learning and artificial intelligence (AI) to support teacher understanding of individual student progress and timely awareness of students who get off track (Goggins & Xing, 2016; Xing et al., 2015, 2018) (see appendix J.2).

B4. Increased Efficiency to Improve Results and Increase Productivity: In each of the five years of the grant, we will implement the most current version of MHS and its components in 5 local classrooms (Continuous Improvement Classrooms, CIC). Each year, starting in the second year, some of the classes will be continuations with teachers and some will include teachers new to MHS. The CICs will allow us to observe game play, discuss practices with the teacher, use focus groups to get student impressions, and identify opportunities for deeper understanding of how gameplay is impacting students. These data will be used to primarily improve the efficiency of implementing MHS and will not be used to change gameplay experience once the field tests have started. There are four main areas where we think we can optimize efficiencies in time and costs for teachers. The first is the planned **myMHS system** for teachers to enroll their classes and have IDs and passwords assigned to each student. Once a teacher has approval for a number of classrooms and students, the teacher will upload a spreadsheet representing student school IDs, email addresses, and passwords (if passwords are known) or have myMHS assign special MHS passwords. The teacher will have access to a master class list on myMHS and each student will be emailed their credentials. myMHS will allow teachers to quickly access the game and assign credentials to all their students, eliminating the need for the school technology coordinator to enroll each student, and saving substantial time. A second initiative is the use and advancement of embedded assessment. Embedded assessment will be described more fully under

Strategy #1 in the following section, but it has the potential to integrate teaching and assessment so that separate and time-consuming post-learning activity assessments will not be needed. The third initiative is to further develop the **Learning Analytics** (**logging and dashboard) system (LA)** (see Appendix J.2) in MHS, which teachers use to monitor students as they progress through MHS. The current MHS has an extensive logging and analytics system designed to capture a broad range of behavioral traces and specific actions related to curriculum and game play progress. The MHS logging system was designed alongside the game itself to log information considered salient for assessing student game and learning progress in real time, and to provide teachers with notifications when student in-game actions indicate they are not progressing. The planned improvements to the logging and analytics system for MHS will include the implementation of (1) new machine learning models, built from the MHS data sets gathered during the development grant field test, and (2) artificial intelligence (AI) experiments for identifying new patterns of engagement. The AI algorithms with the highest accuracy, based on teacher feedback, will be incorporated into the notification system. MHS analytics will also provide teachers with summary dashboards that help make it clear which students are productive, which students need remediation, and what remediation is most likely to be successful. MHS learning analytics will continuously construct, validate, and adapt a set of human interpretable prediction models based on in-game behaviors and match the most appropriate model for each unit to **individual** and **classroom** level patterns of play. The improvements in the dashboard, when combined with embedded assessment described in Strategy 1, will save teachers time in monitoring students, reduce the need for external assessments, and make teachers more effective in supporting students. A fourth approach to increasing efficiency and productivity will be a **tiered approach to the teacher support system.** Teachers in their first year of teaching with

MHS receive an orientation course, a teacher guidebook, access to an online community, knowledgebase, and coaching. In subsequent years, the amount of support needed may be substantially less. For example, a teacher in a second year may no longer need the orientation course and active coach. As eMINTS (a nationally recognized leader in teacher development) designs the TSS, they will design two tiers and we will test the effectiveness of the lower level of staffing and resources needed for the second tier in our two-year teacher study.

C. STRATEGY TO SCALE

C1. Strategies to Scale that Address Prior Barriers: When inviting schools to use MHS, we must overcome three barriers: (1) Would MHS be an effective way of meeting curriculum and assessment objectives? (2) Are teachers interested and able to teach with Game-Based Learning? (3) Could the school run MHS on its available technology? These three questions represent the barriers (along with costs and efficiency) to scaling MHS.

Strategy 1: Continuously improve MHS to meet curriculum and assessment objectives. Since the emergence of the standards movement, research has continually documented pressures experienced by schools and teachers to focus on standards-based content, sometimes to the exclusion of innovative ideas (Barrett-Tatum & Smith, 2018; Pratte, 2001). In short, teachers often feel that they do not have time within a compressed curriculum to introduce new ideas or unproven experiences. For many years, science teachers have cited perceived constraints imposed by curriculum standards as significant impediments to trying new approaches (e.g., Sadler et al., 2006) and this trend has certainly not diminished with the development of NGSS (Castronova & Chernobilsky, 2020). This is why having MHS align with the NGSS is so critical. Indeed, a number of schools cited needing to prepare for state-wide tests as reasons for choosing not to be part of the development grant field test. To make this point in a more positive way, after the field test, one of the school administrators stated the following:

> *After receiving our 8th grade science MAP (state-wide achievement) test scores back and speaking with our students, we feel the program was a successful way of engaging and assisting in students' acquisition of Missouri 8th grade science standards. We would like to get our current 8th graders signed up to be able to use the program.*

Key to the long-term sustainability of MHS is working diligently at improving success for each child and making implementation by teachers as efficient as possible. An advantage of a learning system like MHS is that every lesson learned can potentially be programmed into MHS or the TSS to improve each subsequent edition of MHS for every new teacher and student. There are a number of lessons learned from the field test still to be implemented in MHS, but we also plan to have a small set of classrooms in schools near our development labs that can be used as test beds for trying out improvements and learning more about teacher and student use. These **Continuous Improvement Classrooms (CIC)** will implement MHS annually during the project period. Our team will collect data from the usage as well as through interviews with teachers and students about ways to make MHS efficient for teachers and a good fit for each student.

Another pathway for improving how MHS will support assessment objectives of teachers is **embedded assessment**. Embedded assessment is a way of providing information about student outcomes without external testing, as a way of applying in-game diagnosis to provide alternative pathways for students needing additional or alternative assistance, and as a way of understanding other aspects of learning, such as how students are learning and progressing, not typically available in pre and post testing. Shute and Rahimi (2017) concluded a recent review of computer-based assessments for learning (CBAfL) with a particular call-out for GBL and embedded assessment by stating: "Additional research needs to be conducted on developing systems to deliver valid, reliable, fair and cost-effective CBAfL to accurately measure and improve complex competencies across various disciplines in the near future" (p. 15). We plan to extend the embedded assessment currently operative in MHS by developing alternative pathways for remediation during game play, providing more direct and actionable feedback to teachers and integrating Embedded Assessment with Learning Analytics for more robust models of student learning. More about Embedded Assessment can be found in Appendix J.3.

Strategy 2: Build a system to support teachers as they access, learn about, implement, and advance in their interests and capabilities to teach with MHS. A recent survey of 2000 K–12 teachers shows only 10% feel confident incorporating higher-level technology into student learning (PWC, 2018). The most recent National Educational Technology Plan (2017) summarizes work of researchers in the field of teachers using technology by emphasizing that teachers need continuous, just-in-time support that includes professional development, mentors, and informal collaborations. In fact, more than two-thirds of teachers say they would like more technology in their classrooms, and roughly half say that lack of training is one of the biggest barriers to incorporating technology into their teaching. Bringing this closer to MHS use, the only teacher of the 13 participants in the field test for the development grant, who stated she would not want to use MHS again, explained her decision this way:

> *Oh no, I would not be willing because I think it was a huge deterrent that I don't know how to play. I don't like not knowing what I am doing so I didn't like that feeling.*

The eMINTS National Center, which provides evidence-based professional development programs for K-20 educators, including programs listed in the WWC, will develop the MHS Teacher Support System (TSS). The eMINTS instructional model addresses all aspects of effective instruction required for modern digital age classrooms. Based upon constructivist learning theory, it has four pillars focused on (1) authentic learning, (2) building a community of learners, (3) using high-quality lesson design, and (4) using technology to transform learning. The eMINTS instructional model is supported by over 20 years of external research (Meyers et al., 2016). The TSS will include four initiatives: an orientation course, a teacher guidebook, a teacher community and knowledgebase, and online coaching. These initiatives, in addition to preparing teachers to teach with MHS, will also focus on creating a classroom community and using the positive social nature of game play to help teachers engage students.

In addition, the TSS will facilitate teacher success and maximize student success. TSS will develop and use video to efficiently orient teachers to MHS, to teaching with games, and to using the social environment created by gameplay to improve student's sense of community and engage students who typically are not highly engaged in science learning. **Coaching** from eMINTS facilitators will enable teachers to perform at high levels during the MHS unit and be efficient with their time. The eMINTS coaching model provides virtual support to teachers through planning and reflecting conversations. Building community and community resources for teachers will amplify the support available to teachers with each year of MHS implementation. A tiered TSS will be developed to provide a high level of support in Year 1 of teaching with MHS (Tier 1) and reduced support for cost effectiveness in subsequent years for a given teacher using MHS (Tier 2). More on the teacher support model can be found in App J.4.

Strategy 3: Expand technology platforms to enable more schools to implement MHS on available technology. In 2012, when we were first conceptualizing MHS and began market research on what technology was available in schools to run MHS, all schools we talked to had computer labs of Windows or Macintosh computers. Most of the labs included low performance computers, but we anticipated with Moore's law (computer power doubled and costs halved every 2 years) that by the time we were funded and completed development that school labs would meet the higher performance requirements for playing MHS. As MHS progressed into our development work, we along with the rest of the educational community noticed the dramatic swing away from computer labs and to one-on-one computing with Chromebooks and tablets. The early Chromebooks were primarily web access devices for linking to information on the Internet. After substantial analysis, we determined that the qualities of MHS that we proposed in the development grant could not be implemented via a web-based environment due to the limitations of delivering high-level throughput into classrooms and onto multiple devices. This is still an issue today and probably substantially into the future. At the time, we decided to push through with the implementation of MHS as envisioned with high quality graphics, intense interactions and simulations, analytics, and a feature set that would create the form of student engagement and learning activity that we envisioned as necessary for the sustained and challenging learning, as well as teacher support, needed for meeting NGSS.

Today, with what we have learned from the MHS development work, we have a pathway to optimize MHS for lower capability Windows and Macintosh computers (for which most schools still have labs) and to build a version for app-based devices such as iPads and modern Chromebooks. iPads are capable processing and graphic machines achieving a substantial market share in schools. iPads are the most popular tablet in schools and, in 2017 (the most recent year we have statistics for), sales were up 32% in the U.S. education market (Meaney, 2017). In 2016, Chromebooks became capable of running android apps, but most Chromebooks in schools today are too limited in graphics and processing capabilities for an app providing an immersive, visually rich game like MHS. Building an iPad-based app model for MHS in the mid-phase grant will position MHS to port to an app-based version for Chromebooks in the future. In summary, we plan to expand access to MHS by optimizing a version of MHS to run on lower capability Windows and Macintosh computers and by building an app-based model of MHS to run on iPads. More detail about the work of optimizing/porting and choice of building for iPad rather than Chromebook is discussed in Appendix J.5.

Strategy 4: Develop capacity for sustainability and continued scaling of MHS. MHS requires an infrastructure and maintenance for implementation. The most viable long-term sustainability approach for an innovation like MHS is for it to be a product that schools purchase for use by their teachers and students. Much of our work will be efforts to ensure that the costs to schools are as low as possible and that we create innovative methods for schools, especially schools with high need students who may be less able to afford to pay for use of MHS. Starting in year one of the grant, we will work with the University of Missouri Technology Transfer program to develop a sustainability plan that addresses costs of maintaining the needed infrastructure and maintenance for providing MHS and how to keep those costs low for schools. One key aspect of making MHS available at scale is the need for a system to allow teachers to register their classes, create login IDs for students, and access TSS. We will develop the **myMHS system** (described in B4 about efficiency) for teachers to meet these needs. After the grant period, we envision myMHS as enabling teachers from anywhere in the world to access MHS, implement the needed infrastructure of student ID's needed for the dashboard and student monitoring, and access teacher support.

C2. Broad Dissemination to Support Further Development or Replication: The audiences for MHS dissemination are science teachers and school district leaders, science education and learning technology researchers, and game developers who focus on education. For all of these audiences, we will continue to host a website (MHS.missouri.edu) as well as produce videos for YouTube. We plan to disseminate to academics in science education and learning technologies through traditional journals and conferences, including ICLS (International Conference of the Learning Sciences), NARST (National Association for Research in Science Teaching), AERA (American Education Research Association), and iNACOL (International Association for K12 Online Learning) conferences and *The Journal of Learning Sciences*, *Educational Technology Research and Development*, *Science Education* and the *Journal of Research in Science Teaching*. For teachers and school districts, myMHS will make MHS easily available at a cost recovery basis. We will present about MHS at practitioner-oriented conferences facilitated by organizations such as NSTA (National Science Teacher Association) and ISTE (International Society for Technology in Education). Our partner, MOREnet, has an annual conference as well as connections and communications with all school districts in Missouri. They will provide information to all Missouri schools about MHS. MOREnet is also a member of QUILT, an organization of Research and Education Networks for each of the 50 states. Through this connection, we will be able to disseminate information about MHS to schools in all 50 states and, via myMHS, provide access to MHS to teachers in all 50 states.

For educational game developers, we plan to present at the Connected Learning Summit. We gladly participate in the Office of Educational Technology initiative to openly license all education resources. This initiative will assist us in keeping the cost of MHS to cost recovery for making MHS and services available. We also anticipate that, as part of the development of MHS, we will create assets (software functions such as mechanisms for analytics or representing water systems) that we will make available using open source licensing through the Unity Asset Store.

D. ADEQUACY OF MHS RESOURCES AND MANAGEMENT PLAN

D.1. Capacity to Bring MHS to Scale: Developing, implementing, and evaluating a complex learning game such as MHS requires shared vision and partnership of a number of specialized organizations with established capacities and track records. The School of Information Science and Learning Technologies and the Department of Computer Science at the University of Missouri are the lead organizations responsible to the EIR for grant outcomes. These departments led the i3 development grant for the design, development, implementation, and evaluation of MHS. The four Co-PIs, faculty members Laffey and Goggins and staff members Griffin and Sigoloff, led those i3 efforts as well as forming the Adroit Game Lab to extend and institutionalize the capacity for GBL at MU. The leadership of the four Co-PIs will manage and coordinate all partner activity as well as directly lead the development and implementation efforts of Adroit. Adroit will be responsible for (1) development and testing work to optimize MHS for lower capacity Win/Mac computers and porting to iPads, (2) continuous improvement of MHS, (3) recruitment and implementation of MHS during the field testing, and (4) planning for beyond the grant period.

The **eMINTS National Center at the University of Missouri** will develop and implement the TSS. eMINTS has over 18 years of experience helping teachers combine technological, pedagogical, and content knowledge to improve classroom instruction. eMINTS has managed two federal i3 validation grants awarded in 2010 and 2015. The eMINTS Comprehensive Professional Development Program is one of the few programs with data to support the chain of evidence from delivery of a specific technology professional development to changing teacher practice and to positive impacts on student achievement (Meyers et al., 2016). A recent randomized control trial of the program found significant, positive teacher outcomes as well as student outcomes that meet What Works Clearinghouse standards without reservations.

The **Missouri Research and Education Network (MOREnet)** will develop and implement myMHS for teacher registration and provide the infrastructure for the online system of LA. MOREnet is a department within the University of Missouri System that provides Internet connectivity and essential technical services to more than 700 of Missouri's public sector entities, including K-12 schools, colleges and universities, public libraries, health care, state agencies, local government, and non-profit organizations. MOREnet is a member-funded organization based on a cost-recovery model for service offerings. As a member of The Quilt, a national community of research and education networks, MOREnet has connected MHS with OneNet, the education network for Oklahoma and UEN, the education network for Utah.

Abt Associates. Abt is well regarded for its rigorous approach to solving complex challenges and has led numerous high-profile, innovative studies and rigorous impact and implementation evaluations for multiple agencies in the federal government, states, and foundations, including three current EIR early and mid-phase grant evaluations. The Abt evaluation team is well-versed in the What Works Clearinghouse (WWC) evidence standards; all team members are certified WWC reviewers. The Abt team offers a depth of knowledge about evaluations and the evidence requirements for EIR grant evaluations. The evaluation project director, Michelle Blocklin, is a Senior Associate and a WWC certified reviewer.

D.2. Extent to Which Costs are Reasonable: Costs of developing commercial adventure-narrative games often exceed $50 million. Although MHS does not match those games for the full level of graphics, animation, and rapid action, based on reports from students, we come close enough that students compare MHS to those games and talk about MHS as being cool to be able to play video games in school. We believe the costs of development so far and those described in this proposal represent a cost-effective methodology for making science learning both fun and effective. In addition to the costs of developing MHS, optimizing, and porting, the other significant costs (unrelated to evaluation) include developing infrastructure and teacher support systems which will leverage the benefits of MHS and sustain MHS at a lowest possible cost.

As noted in the significance section, MHS is the only science game with outcomes that meet WWC standards. MHS also meets important NGSS outcomes desired in schools, and while games hold great promise for engaging students in important learning outcomes, most games in schools are short-form games used to simply introduce a topic or provide drill and practice (Takeuchi & Vaala, 2014). In addition to leading to specific student outcomes, MHS can be a model for how to use the powerful adventure-narrative type game for educational objectives. An added benefit of funding the current project is that with the new open license requirements of the U.S. Department of Education, the current grant will transfer the license from the current IP owners to an open-source license, further enabling access to schools at the lowest possible costs.

D.3. Continuing MHS Beyond the Grant: Sustaining MHS beyond the grant period is dependent upon it being a cost-effective way to achieve standards-aligned learning outcomes for students. The significance section makes the case for MHS being effective at delivering NGSS learning objectives, which are in demand in schools. However, will schools be able to pay for the infrastructures of teacher support, access, and networking needed to run MHS, and continued maintenance required for all software systems? The University of Missouri Technology Transfer office will provide assistance in setting up Adroit to provide MHS to schools (see letter of support). MOREnet operates on a cost-recovery basis, so the costs of implementing MHS will be kept as low as possible. We are developing and testing a tiered TSS so that while there will be costs to support teachers as they initially implement MHS in order to achieve high student outcomes, those costs can be substantially reduced in subsequent years of implementation. In addition, in order to assure that high need students in urban and rural schools can benefit from MHS, the Advancement Office at MU will set up and solicit funds for an "MHS Scholarship" program for schools that need help in funding MHS for their students (see letters of support). A strength of technology tools like MHS is that while the costs of development are high, much of the costs for implementation can scale rapidly such that providing MHS to 10,000 students might cost $8 per student, but providing MHS to 50,000 students might reduce the costs to $3 per student. We will also look for other innovative ways to help schools pay for MHS, including using the community of teachers to help other teachers.

D.4. A Management Plan for Responsibilities, Timelines, and Milestones

Exhibit 2: Milestones, Responsible Partner, and Timeline for Each Objective (smaller font indicates follow-up and refinement)

		Project Year (Jan 1 to December 31)				
Milestones	**Responsible**	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Obj 1: Increase the number of schools that can use MHS by expanding the platforms schools can use to implement MHS						
MS 1.1	SISLT, Adroit	X	X	X	x	x
MS 1.2	SISLT, Adroit	X	X	X	x	x
Obj 2: Plan for scale-up by improving efficiency in accessing MHS and by disseminating information for further development, usage, and replication						
MS 2.1	MOREnet	X	X	X	x	x
MS 2.2	SISLT, Adroit, MOREnet			X		
MS 2.3	SISLT, Adroit, MOREnet, eMINTS			X	X	X
Obj 3: Improve the effectiveness and efficiency of using MHS for teachers by enhancing the Teacher Support System (TSS)						
MS 3.1	eMINTS	X	X	X	x	x
MS 3.2	eMINTS			X		
Obj 4: Continuous improvement of MHS throughout the project						
MS 4.1	SISLT, Adroit	X	X	X	X	X
MS 4.2	Computer Science	X	X	X	x	x
MS 4.3	SISLT, Adroit	X	X	X	x	x
MS 4.4	SISLT, CS, Adorit, eMINTS, MOREnet	X	X	X	X	X
Obj 5 Impact Study of MHS with (60) teachers in Years 3, 4, & 5 of grant						
MS 5	Abt			X	X	X

Obj 6 Understand and explore implementation issues of MHS						
MS 6.1	SISLT, Adroit			X	X	X
MS 6.2	SISLT, Adroit, eMINTS			X	X	X
MS 6.3	SISLT, Adroit			X	X	X
Obj 7 Understand, explore, and plan for scale-up of MHS						
MS 7.1	Abt, SISLT, Adroit			X	X	X
MS 7.2	Abt, SISLT, Adroit			X	X	X

E. PROJECT EVALUATION

Abt will conduct an independent evaluation of MHS, building on the previous efficacy evaluation (Reeves et al., 2020). In the proposed study, we will conduct a randomized controlled trial--the most rigorous possible design--to test whether MHS achieves the same effectiveness when implemented in a broader array of settings, with more heterogeneous populations of middle school students, and using delivery mechanisms that will improve the long-term scalability of the program. The evaluation is designed to assess both the impact and success of implementation of the two versions of MHS—mac/win and app-based. In addition, the evaluation will test the success of MHS with reduced teacher support (Tier 2 TSS) as a key scale-up component. Accordingly, the evaluation will answer ten important research questions (Exhibit 3).

Exhibit 3. Evaluation Research Questions

Impact Study Research Questions
1. What is the impact of MHS (mac/win or app-based version) compared to business-as-usual on middle school students' water systems knowledge, when teachers are in their first year of implementing the intervention? a. What is the impact of the mac/win version of MHS? b. What is the impact of the app-based version of MHS?
2. What is the impact of MHS (mac/win or app-based version) compared to business-as-usual on middle school students' scientific argumentation, when teachers are in their first year of implementing the intervention? a. What is the impact of the mac/win version of MHS? b. What is the impact of the app-based version of MHS?
Scale-Up Study Research Questions
1. What is the cost-effectiveness of each of the versions of MHS?
2. Does fidelity vary for teachers in their first year of implementation compared to teachers in their second year of implementation with a lower-cost teacher support system?
3. How does the impact of MHS on students with teachers in their second year of teaching MHS with a lower-cost teacher support system compare to the impact of MHS on students with teachers in their first year of teaching MHS?

4.	What are teachers' assessments of the quality of the low-cost support for implementing MHS in the second year of implementation compared to the support offered in the first year of implementation?
5.	What barriers/challenges to successful scale-up did the project encounter that would need to be addressed at the next level of scaling up?
Implementation Study Research Questions	
1.	What is the level of fidelity of implementation of MHS in middle school classes?
2.	At what levels and in what ways do students engage in MHS?
3.	What are teacher attitudes toward the value of MHS as an approach to teaching water systems and scientific argumentation?

E1. Evaluation Methods Designed to Produce Evidence that will Meet WWC Evidence Standards without Reservations:

Design. The impact study will use a blocked cluster randomized controlled trial design, with classes within teachers as the unit of assignment. This experimental design is eligible to receive the highest WWC evidence rating of Meets Standards without Reservations. The impact study will be conducted in two cohorts of schools over two school years (SY 2024-25 and SY 2025-26). In each school, for each participating teacher, one class section will be randomly selected to receive the business-as-usual (BAU) water science curriculum, and the remaining classes will receive the MHS curriculum. Teachers will implement the MHS and BAU water systems units at the same time during the school year to ensure there are no time confounds. Given that the MHS intervention is delivered via a computer or tablet, all MHS students receive a unique login that will not be available to BAU students, teachers will be instructed to not share approaches across classes, and since the treatment period is relatively short, the risk of contamination between treatment and control classes is very low. The intervention will be tested by combining results from the two cohorts. The design will test the impact of MHS across the two modes of delivery and will also look separately at the effects of the mac/win and app-based versions.

Recruitment, Sample, and Power. Interested middle school science teachers will be recruited to participate in the evaluation. Recruitment will be done in Midwest states whose state education network provider (such as MOREnet, UEN, & Onenet) will be recruited by MOREnet through Quilt. For these states, we will also use state teacher associations and focus on recruiting teachers from districts where the student population includes a majority of students considered high need, defined as receiving free or reduced-price school lunch and scoring below proficient on state tests. Once teachers are recruited, we will recruit their schools for the study, completing research applications and generating MOUs with districts and schools. To be eligible for the study, schools will have to have sufficient mac/win or iPads to fully implement MHS. Further, the middle school science teacher must teach at least two class periods within the specific grade that covers water science, i.e., at least two class periods for 6th grade, 7th grade, or 8th grade, to support randomization of classes (of the same grade) within teachers. Teachers and their technology coordinators will be compensated for their effort and extra time with a stipend ($1000 per teacher; $250 per technology coordinator).

Across an estimated 20 districts and 40 schools, the study will include a sample of 58 teachers over the course of two years: 29 science teachers who use computers that support the mac/win version of MHS and 29 science teachers who use tablets that support the iPad version of MHS. All of the recruited teachers' science classes will participate in the study for a total of 203 classes (assuming an average of 3.5 classes per teacher with one randomized to control and the remaining to treatment). All students in these classes with parental permission will participate in the study, for a total of 4,669 students (assuming an average of 23 students per class). These sample sizes shown in Exhibit 4 mean that the study can detect an effect size of .20 for each of the mac/win and the app-based versions of MHS (see Appendix J.6). An effect size of .20 is comparable to the effect size found in the previous efficacy study. Pooling across the mac/win and app-based samples will allow us to detect an overall effect size of .14[1].

Teachers in the first cohort will also be asked to participate for a second year with reduced support (Tier 2 TSS), and for the anticipated 75% who agree, we will re-randomize their science classes to MHS or BAU (as in Year 1), and students in these classes who receive parental consent (prior to random assignment) will participate in the study.

Exhibit 4. Projected Sample Sizes for Analyses after 1 and 2 Years of Implementation

		Cohort 1 (Teachers begin in SY2024-25)			**Cohort 2 (Teachers begin in SY2025-26)**		**Total Sample**
		T	**C**		**T**	**C**	
Sample for impacts at end of teachers' 1st year of implementation	Teachers[a]	29		Teachers[a]	29		**58**
	Classes[b] (cohort 1a)	72.5	29	Classes[b] (cohort 2a)	72.5	29	**203**
	Students[c] (cohort 1a)	1667.5[d]	667[d]	Students[c] (cohort 2a)	1667.5[e]	667[e]	**4669**
Sample for impacts at end of teachers' 2nd year of implementation	Teachers[a]	22					**22**
	Classes[b] (cohort 1b)	55	22				**77**
	Students[c] (cohort 1b)	1265[e]	506[e]				**1771**

[a] Half of the teachers in each cohort will be recruited to use the mac/win version and half will be recruited to use the app-based version of MHS.
[b] We assume an average of 3.5 classes per teacher with 1 assigned to control, the remainder to treatment.
[c] We estimate 23 students per class.
[d] Student outcomes will be assessed in SY2024-25.
[e] Student outcomes will be assessed in SY2025-26.

Strategies to Exclude Joiners and Minimize Attrition. Our randomization strategy allows us to exclude joiners from the sample and minimize attrition in alignment with the WWC 4.1 standards. Prior to random assignment, the project team will work with schools to identify all class sections of the recruited teachers and obtain rosters of students assigned to each class

[1] Power analyses were conducted using *PowerUp!* (Dong & Maynard, 2013) with the following assumptions based on the previous efficacy study: ICC = .10, 73% of classes assigned to treatment, level 1 and level 2 R^2 = .3; 23 students per class; 3.5 classes per teacher.

section. Following school district and MU IRB requirements, the project team will obtain parental consent for student participation in the evaluation. All students with parental consent will be included in the study sample. After the sample of classes and consented students is identified, Abt will randomly assign classes within teachers as described above. Identifying the sample of students prior to random assignment will allow the study team to exclude any joiners from the sample, thereby eliminating what the WWC considers to be a threat to the internal validity of the study and its ability to meet standards without reservations. Another threat to the internal validity of a randomized controlled trial is attrition. In this study, the pre-test, intervention or BAU unit, and post-test for each cohort will all take place over the course of a month or less. In addition, we will coordinate with teachers to track, monitor, and remind students to complete surveys. Given the short time period between pre- and post-test combined with our approach to achieving high response rates, we expect attrition to be minimal. In the event that attrition or differential attrition is much higher than expected, our collection of baseline data will allow for the assessment of baseline equivalence and the inclusion of baseline covariates in the analytic model if adjustment is needed.

Data Collection of Valid and Reliable Outcome Measures: To meet WWC evidence standards without reservations, studies must also include valid and reliable outcome measures. The key long term study outcomes are water systems knowledge and scientific argumentation, both of which align with Next Generation Science Standards (NCCS) and state-based science standards. The Water Systems Assessment (WSA) was developed and validated for the previous efficacy study. The instrument comprises 23 multiple-choice items ($\alpha = 0.72$) that address multiple dimensions of Earth water systems in the middle school learning standards in the NGGS, including watersheds, surface water, groundwater, and water cycle processes (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. The Argumentation Assessment (AA) was developed in prior work by a team of researchers studying Argumentation Driven Inquiry (ADI) independent of MHS (Osborne et al., 2013, 2014; Grooms et al., 2014). The AA is made up of 12 multiple-choice items ($\alpha = 0.60$) related to a water-themed scenario. There are three item clusters that challenge students to identify critical components of an argument's structure, align evidence to a given claim, and critique arguments (Sadler et al., 2019; Wulff, 2019). These outcome measures both meet WWC standards for reliability[2] and validity (see Appendix J.7 for the questions included in these measures).

In addition to these outcome measures, we will also measure student engagement as an intermediate outcome, using the engagement subscale of the Panorama Student Survey (https://www.panoramaed.com/student-survey-questions). The 5 item subscale measures students' attentiveness and investment in a class. Psychometric studies have established reliability ($\alpha > .70$) as well as structural and convergent/discriminant validity of this measure (https://go.panoramaed.com/hubfs/Panorama_January2019%20/Docs/validity-brief.pdf), meeting WWC standards. Pre-test student surveys will ask for basic demographic information (gender, age, race/ethnicity) as well as a self-rating of how they are doing in their current science class.

Teachers will teach the water systems unit at the same time in their BAU and MHS class sections. The evaluation team will coordinate with teachers to administer 40-minute web-based student assessments (including WSA, AA, and student engagement measures) during the class period on the day prior and the day following the implementation of MHS and the BAU units.

[2] According to the WWC, reliability of outcome measures is established with internal consistency (α) of .50 or higher (WWC Standards Handbook version 4.1, page 83).

Accordingly, data collection procedures and timing will be the same for treatment and control students. The evaluation team will also pilot outcome measures and data collection procedures with the pilot cohort of eight teachers and their students during Spring 2024.

Analysis. The approach to the impact analysis aligns with the randomized block design stratified by teacher, using a 2-level regression model where students (level 1) are nested within classrooms (level 2) and includes a series of dummy variables to represent randomization blocks (teachers)[3]. This model adjusts standard errors to account for the dependency among students within classrooms, thereby avoiding the overestimation of statistical significance of the impact estimate. Impacts will be estimated using an intent-to-treat analysis to compare treatment and control group means on water systems knowledge, scientific argumentation, and student engagement. Baseline measures of the outcomes will be included in the analytic model to improve precision and increase power. WSA baseline scores will also be included for the argumentation outcome to account for prior content knowledge as this may affect a student's argumentation skills. Additional student covariates to be included in the models are gender, age at baseline, race/ethnicity, and self-rating of science achievement.

Analyses to assess the overall impact of MHS will combine data from students of both cohorts of teachers from teachers' first (cohort 1a and cohort 2a) and second years of implementation (cohort 1b) and across both mac/win and app-based versions of MHS. We will also conduct separate analyses for the mac/win version and the app-based version, following the same analytic approach. To test for whether student impacts vary based on whether teachers are in their first year or in their second year of teaching MHS with the lower cost teacher support system, we will include all students in the analysis and include a treatment*implementation year

[3] These randomization blocks will account for different probabilities of assignment across teachers.

interaction term in the model. See Appendix J.8 for the statistical model describing the analytic approach.

We will not impute outcome data. For each outcome, we will calculate overall and differential attrition as the students with missing data on the outcome divided by the total number of students in the baseline sample. We expect to have very high response rates for the baseline surveys which will provide pre-test and covariate data. In the event of missing covariate data, we will use the dummy variable method (Puma et al., 2009) implemented in accordance with WWC evidence standards. In addition, we will only rely on one outcome measure per outcome domain and will not need to adjust for multiple comparisons.

E2. Evaluation will Provide Guidance about Effective Strategies Suitable for Replication or Testing in Other Settings: The scale up study questions are shown in Exhibit 3. To answer the question on *cost-effectiveness*, we will also explore the feasibility of using an available cost-effectiveness tool such as the IES Cost Analysis Starter Kit or the CostOut Tool[4], which would allow for a rigorous assessment and comprehensive understanding of the cost-effectiveness of MHS if the necessary data elements are available. At a minimum, the project team will collect, track, and share data on the costs of implementing MHS with Abt. These costs will include costs per teacher and costs per student. These data are reported each year to OESE as part of Annual Program Reporting. The grantee will remove costs of the evaluation from the overall annual costs and provide the number to Abt, who will then link the student effect size to the costs.

We will also examine the extent to which the study has met its other scale-up goals to support the replication and testing of MHS in other settings. To complement the impact study's test of the expanded MHS technology platforms (mac/win and iPad), the effectiveness of Tier 2

[4] https://ies.ed.gov/seer/cost_analysis.asp

TSS (lower-cost support system) will be examined in multiple ways to thoroughly assess the MHS scale-up mechanisms. First, we will look at the fidelity of teacher implementation (described in E3 below) when teachers are in their first compared to their second year of MHS implementation, when teachers are given a lower level of support (Tier 2 TSS), which is likely to be the case in the real world during further scale-up. Second, we will test for the difference in impact between full teacher support in their first year and the lower-cost support system in their second year (described in E1 above). Third, in order to further understand the adequacy of the lower-cost teacher support system scale-up mechanism, we will also conduct interviews with all teachers in their first and second year of implementing MHS. These interviews will also allow us to understand teachers' perception of MHS (both mac/win and iPad) as an effective, efficient, and desirable way to teach water systems and scientific argumentation, which is also key to a nationwide roll-out of MHS. Interviews will also uncover any barriers to implementing the scaled-up version of MHS that will need to be addressed prior to further scale-up. We will conduct qualitative analysis of interview data in NVivo to identify key themes related to each of our research questions.

To further address our final scale-up question on lessons learned across the full project effort to identify additional barriers and challenges to the next level of scaling up, each year, with the assistance of the MU technology transfer and business development offices, the Co-PIs will examine the opportunities and barriers to sustainability. In Years 1, 3, & 5, experts from business development will be consulted for additional perspectives. Aspects of the plan will include developing an enterprise to sustain and provide infrastructure for implementing MHS, including Adroit, MOREnet, and eMINTS. Aspects of the plan will also address most appropriate means for disseminating information about MHS and ways of making MHS available via the Open

Source License. Given the impact on schools, financially, and with the need for social distancing and technology solutions, we expect the next several years and into the future to be challenging yet filled with opportunities to meet real needs for educational solutions. These scaling and sustainability planning efforts will lead to an annual report/guide for project action, but will also lead to an implementation of a plan to make MHS available to schools in the Fall of 2026. We will monitor these sustainability and scaling efforts and examine the number of teachers who sign up to use MHS in the 2026-27 school year (after the impact evaluation period) as a proxy for their effectiveness.

E3. Clear Articulation of the Key Components, Mediators, and Outcomes, and a Measurable Threshold for Acceptable Implementation: In addition to the impact study and scale-up study described above, a comprehensive implementation study will address the research questions listed in Exhibit 4 above. The logic model (see Appendix G) clearly articulates how the *key components* (or inputs) of MHS (game play components, teacher support system, expanded platforms, and capacity for scale-up including myMHS access) produce outputs, which lead to teacher and student implementation of/participation in MHS. These *mediators* are theorized to increase student engagement, as an intermediate student outcome by providing a fun, game-based experience that will be especially engaging for students not-typically highly engaged in science learning (as shown in the earlier teacher quotations). The impact study, described above, will assess the impact of MHS on student engagement as an intermediate outcome, and the implementation study will also assess student engagement in MHS in two ways: (1) an implementation survey of MHS students mid-way through their participation in MHS based on modifications to the Game Engagement Questionnaire (Brockmyer et al., 2009), and (2) an embedded assessment of engagement based on constructs of perseverance, off task behavior, and gaming the system developed through analytics (Rowe et al., 2011) as well as through study efforts in the CIC classrooms. We will conduct descriptive analyses of these data to understand the level of engagement of MHS students during the intervention. The impact study, described above, is designed to assess these long-term outcomes as articulated in the MHS logic model.

One of the key implementation study research questions addresses the fidelity of implementation of MHS. We will assess the fidelity of implementation of the key components of MHS as shown in the logic model, using an adapted version of the fidelity matrix developed during the efficacy study (see Appendix J.9), which met EIR criteria. The revised fidelity matrix will include all key components of the current project, incorporating the full TSS as well as the Tier 2 (less intensive) TSS, which is a new scale-up mechanism of this implementation of MHS. The measure will include the indicators and *thresholds for acceptable implementation* at both the unit of implementation and across the full sample, which will ultimately allow us to conclude whether each key component of MHS is implemented with fidelity across the full sample in each implementation year (with full and Tier 2 TSS).

To answer the third implementation study research question, we will assess teacher attitudes about using new technology by asking teachers three questions on the key constructs assessed for the Technology Acceptance Model (Scherer & Teo, 2019). The TAM explains teachers' intentions to use technology by their attitudes toward technology, which are in turn predicted by their beliefs about technology. We will use interview questions to elicit teacher attitudes and beliefs about the constructs and conduct a qualitative analysis of the responses to extract key themes and the valence of attitudes toward using MHS to teach science and engage students.

REFERENCES

Barab, S. A., Sadler, T. D., Heiselt, C., Hickey, D., & Zuiker, S. (2007). Relating narrative, inquiry, and inscriptions: Supporting consequential play. *Journal of Science Education and Technology, 16*(1), 59-82.

Barab, S. A., Gresfali, M. S., & Arici, A. (2009). Why educators should care about games. *Educational Leadership, 67*(1), 76–80.

Barrett-Tatum, J., & Smith, J. M. (2018). Questioning reform in the standards movement: Professional development and implementation of common core across the rural South. *Teachers and Teaching*, *24*(4), 384–412. https://doi.org/10.1080/13540602.2017.1401534

Bathgate, M., & Schunn, C. (2016). Disentangling intensity from breadth of science interest: What predicts learning behaviors. *Instructional Science, 44*, 423-440.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education, 94*(5), 765-793.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Riberio, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Bryan, R. R., Glynn, S. M., & Kittleson, J. M. (2011). Motivation, achievement, and advanced placement intent of high school students learning science. *Science Education, 95*(6), 1049–1065. doi:10.1002/sce.20462.

Castronova, M., & Chernobilsky, E. (2020). Teachers' pedagogical reflections on the next generation science standards. *Journal of Science Teacher Education*, *31*(4), 401–413. https://doi.org/10.1080/1046560X.2019.1710387

Cavagnetto, A. R. (2010). Argument to foster scientific literacy. A review of argument interventions in K-12 science contexts. *Review of Educational Research, 80*(3), 336- 371.

Clark, D., Tanner-Smith, E., & Killingsworth, S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122. DOI: 10.3102/0034654315582065

Clarke, J., Dede, C., Ketelhut, D. J., Nelson, B., & Bowman, C. (2006). A Design-based research strategy to promote scalability for educational innovations. *Educational Technology, 46(3)*, 27-36.

Connolly, T., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661-686.

Covitt, B. (2013). Reasoning tools for understanding water systems. Retrieved from http://www.umt.edu/watertools/default.aspx

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.

Dede, C., Grotzer, T. A., Kamarainen, A., Metcalf, S., & Tutwiler, M. S. (2012). EcoMOBILE: Blending virtual and augmented realities for learning ecosystems science and complex causality. *Journal of Immersive Education*. http://JiED.org/1/1/2

Dickerson, D., & Dawkins, K. (2004). Eighth grade students' understandings of groundwater. *Journal of Geoscience Education, 52*, 178–181.

Dong, N., & Maynard, R. A. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, *6*(1), 24-67. doi: 10.1080/19345747.2012.673143

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 46*, 606–609.

Eastwood, J. L., & Sadler, T. D. (2013). Teachers' implementation of a game-based biotechnology curriculum. *Computers & Education*, *66*, 11–24. https://doi.org/10.1016/j.compedu.2013.02.003

Endreny, A. H. (2010). Urban 5th graders conceptions during a place-based inquiry unit on watersheds. *Journal of Research in Science Teaching, 47*, 501–517.

Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of Econometrics*, *2*, 775-826.

Entertainment Software Association, (2019), The 2019 Essential Facts about the Computer and Game Industry. A trade report of the ESA. https://www.theesa.com/wp-content/uploads/2019/05/2019-Essential-Facts-About-the-Computer-and-Video-Game-Industry.pdf

Erduran, S., Simon, S., & Osborne, J. (2004). TAPing into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88*, 915-933.

Erduran, S., & Jiménex-Aleixandre, M. P. (2008). *Argumentation in science education: Perspectives from classroom-based research*. Springer.

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge - LAK '15*, 146–150. https://doi.org/10.1145/2723576.2723589

Fiaidhi, J. (2014). The next step for learning analytics. *IT Professional*, *16*(5), 4–8. https://doi.org/10.1109/MITP.2014.78

Fox, C. M., & Brockmyer, J. H. (2013). The development of the game engagement questionnaire: A measure of engagement in video game playing: Response to reviews. *Interacting with Computers, 25*, 290-293.

Ge, X., & Ifenthaler, D. (2018). Designing engaging educational games and assessing engagement in game-based learning. In *Gamification in education: Breakthroughs in research and practice* (pp. 1-19). IGI Global.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy* (pp. 2, 14, 203-210). Palgrave Macmillan.

GlassLab: About SimCityEDU. (2013). http://www.simcityedu.org/about/.

Goggins, S. P., Mascaro, C., & Valetto, G. (2013a). Group informatics: A methodological approach and ontology for sociotechnical group research.*Journal of the American Society for Information Science and Technology*, *64*(3), 516-539.

Goggins, S. P., Xing, W., Chen, X., Chen, B., & Wadholm, B. (2015). Learning analytics at "small" scale: Exploring a complexity-grounded model for assessment automation. *Journal of Universal Computer Science*, *21*(1), 66–92.

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education*, *94*, 241–251.

Goggins, S., Galyen, K., Petakovic, E., & Laffey, J. (2016). Connecting performance to social structure and pedagogy as a pathway to scaling learning analytics in MOOCs: An exploratory study. *Journal of Computer Assisted Learning*, *32*(3), 244–266.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood to late adolescence: A longitudinal study. *Journal of Educational Psychology, 82*, 525–538.

Griffin, J., Kim, S. M., Sigoloff, J., Sadler, T. D., & Laffey, J. (2016, August). *Designing scientific argumentation into Mission HydroSci.* Proceedings of the Games + Learning + Society Conference. Madison, WI.

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, 9*, 1412-1433.

Grooms, J., Sampson, V., & Enderle, P. (2018). How concept familiarity and experience with scientific argumentation are related to the way groups participate in an episode of argumentation. *Journal of Research in Science Teaching*, *55*(9), 1264–1286. https://doi.org/10.1002/tea.21451

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.

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching, 49*(7), 843-868.

Haag, S., & Megowan, C. (2015). Next Generation Science Standards: A national mixed-methods study on teacher readiness. *School Science and Mathematics, 115*, 416-426.

Harri, K., Sithole, A., & Kibirige, J. (2017). A needs assessment for the adoption of Next Generation Science Standards (NGSS) in K-12 education in the United States. *Journal of Education and Training Studies, 5*. http://jets.redfame.com

Hatch, J. A. (2002). *Doing qualitative research in education settings*. SUNY Press.

Hawkey, R., & Clay, J. (1998). Expectations of secondary science: Realisation and retrospect. *School Science Review, 79*(289), 81–83.

Hogan, K., & Corey, C. (2001). Viewing classrooms as cultural contexts for fostering scientific literacy. *Anthropology & Education Quarterly, 32*(2), 214-243.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*(5958), 1410–1412.

Introne, J., & Goggins, S. (2019). Advice reification, learning, and emergent collective intelligence in online health support communities. *Computers in Human Behavior*, *Accepted*.

Introne, J., Erickson, I., Semaan, B., & Goggins, S. (2019). Designing sustainable online support: Examining the effects of design change in forty-nine online health support communities. *JASIS&T*, *70*(8).

Introne, J., Semaan, B., & Goggins, S. (2016). A sociotechnical mechanism for online support provision. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3559–3571. ACM.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education, 86*(3), 314- 342.

Laffey, J., Stichter, J., Galyen, K., Wang, X., Ding, N., Babiuch, R., & Griffin, J. (2013). iSocial: Collaborative distance education for special needs. *International Conference of Computer-Supported Collaborative Learning (CSCL 2013).* Madison, WI. 299-300.

Laffey, J. M., Griffin, J., Sigoloff, J., Sadler, T., Goggins, Womack, A. J., Wulff, E., & Lander, S. (2019). Mission HydroSci: Meeting learning standards through gameplay. In K. Lund, G. P. Niccolai, E. Lavoue, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.) *A wide lens: Combining embodied, enactive, extended, and embedded learning in collaborative setting, 13th International Conference on Computer Supported Collaborative Learning* (pp. 1017-1020). Lyon, France: International Society of the Learning Sciences.

Laffey, J., Sadler, T., Goggins, S., Griffin, J., & Babiuch, R. (2016). Mission HydroSci: Distance learning through game-based 3D virtual learning environments. In D. Russell & J. Laffey (Eds.), *Handbook of research on gaming trends in P-12 education* (pp. 421-441). IGI Global. doi:10.4018/978-1-4666-9629-7

Laffey, J., Reid, D., Hong, R-Y., Galyen, K., Xie, X., & Keuker, D. (2011). Activity notification: Enhancing the social nature of online learning. *2011 Annual Conference of ED-MEDIA*, Lisbon, Portugal.

Lamb, R. L., Annetta, L., Vallett, D. B., & Sadler, T. D. (2014). Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames. *Computers & Education*, *70*, 92–104. https://doi.org/10.1016/j.compedu.2013.08.008

Lee, C. S., Hayes, K. N., Seitz, J., DiStefano, R., & O'Connor, D. (2016). Understanding motivational structures that differentially predict engagement and achievement in middle school science. *International Journal of Science Education, 38*, 192-215.

List of most expensive video games to develop. (March, 2020). Retrieved on June 1, 2020. From https://en.wikipedia.org/wiki/List_of_most_expensive_video_games_to_develop.

Marklund, B. B., & Taylor, A.-S. A. (2016). Educational games in practice: The challenges involved in conducting a game-based curriculum. *Electronic Journal of E-Learning*, *14*(2), 122–135.

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research.* London: Routeldge.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*(2), 153-191.

McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. In M. C. Lovett & P. Shah (Eds.), *Thinking with data*: *The Proceedings of the 33rd Carnegie Symposium on Cognition* (pp. 233-265). Mahwah, NJ: Erlbaum.

Meaney, K. (2017). Apple's iPad is Most Popular Tablet in Schools. Retrieved May, 23, 2020. https://www.simbainformation.com/Content/Blog/2017/08/07/Apples-iPad-is-Most-Popular-Tablet-in-Schools

Medina, R., & Suthers, D. (2012). Inscriptions becoming representations in representational practices. *Journal of the Learning Sciences*, 1–37. doi:10.1080/10508406.2012.737390

Metcalf, S. J., Clarke, J., & Dede, C. (2009). Virtual worlds for education: River City and EcoMUVE. In *MiT6 International Conference,* 1-6.

Meyers, C. V., Molefe, A., Brandt, W. C., Zhu, B., & Dhillon, S. (2016). Impact results of the eMINTS professional development validation study. *Educational Evaluation and Policy Analysis*, *38*(3), 455-476.

Meyers, C., Molefe, A., Dhillon, S., & Zhu, B. (2015). The impact of eMINTS professional development on teacher instruction and student achievement. Year 3 report. Washington, DC: American Institutes for Research.

Mislevy, R., Almond, R., & Lukas, J. (2004). *A brief introduction to evidence-centered design*. Center for Research on Evaluation, Standards, and Student Testing.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by state*s. The National Academies Press.

National Research Council (NRC). (2000). Inquiry in the national science education standards: A guide for teaching and learning. National Academy Press.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, *21*(5), 553-576.

NETP. (2017). Reimagining the role of technology in education. Retrieved May, 23, 2020. https://tech.ed.gov/files/2017/01/NETP17.pdf

NRC. (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education, 95*(4), 627-638.

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.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079.

Plass, J. L., Mayer, R. E., & Homer, B. D. (Eds.). (2020). *Handbook of game-based learning*. MIT Press.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*(1), 61–86.

Pratte, R. (2001). Standards movement, accountability, and responsibility. *Journal of Thought*, *36*(3), 35–44. JSTOR.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). What to do when data are missing in group randomized controlled trials (NCEE 2009-0049). Washington DC. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

PWC (2018). Are we preparing our kids for the jobs of tomorrow? Retrieved May, 23, 2020. https://www.pwc.com/us/en/about-us/corporate-responsibility/assets/pwc-are-we-preparing-our-kids-for-the-jobs-of-tomorrow.pdf

Qian, M., & Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, *63*, 50-58.

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

Remillard, J. T., & Bryans, M. B. (2004). Teachers' orientations toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education*, *35*(5), 352–388. https://doi.org/10.2307/30034820

Reeves. T. C., Romine, W., Laffey, J., Sadler, T., & Goggins, S. (2020). Distance learning through game-based 3D virtual learning environments: Mission HydroSci (External Evaluation Report 1). University of Georgia. (ERIC Document Reproduction Service No. TBA).

Rowe, J., Shores, L., Mott, B., & Lester, J. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, *21*(1-2), 115-133.

Sadler, T. D., Amirshokoohi, A., Kazempour, M., & Allspaw, K. M. (2006). Socioscience and ethics in science classrooms: Teacher perspectives and strategies. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *43*(4), 353–376.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching, 41*(5), 513-536.

Sadler, T. D., Romine, W. L., Stuart, P. E., & Merle-Johnson, D. (2013). Game-based curricula in biology classes. Differential effects among varying academic levels. *Journal of Research in Science Teaching, 50*(4), 479-499.

Sadler, T. D., Romine, W. L., Menon, D., Ferdig, R. E., & Annetta, L. (2015). Learning biology through innovative curricula: A comparison of game- and nongame-based approaches. *Science Education*.

Sadler, T. D., Romine, W. L., Menon, D.*, Ferdig, R. E., & Annetta, L. (2015). Learning biology through innovative curricula: A comparison of game- and nongame-based approaches. *Science Education, 99*, 696-720. DOI: 10.1002/sce.21171

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.

Sadler, T. D., & Donnelly, L. A. (2006). Socioscientific argumentation: The effects of content knowledge and morality. *International Journal of Science Education, 28*(12), 1463-1488.

Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education, 128*, 13-35.

Shah, M., & Foster, A. (2014). Undertaking an ecological approach to advance game-based learning: A case study. *Journal of Educational Technology & Society*, *17*(1), 29–41.

Shepardson, D. P., Wee, B., Priddy, M., Schellenberger, L., & Harbor, J. (2007). What is a watershed? Implications of student conceptions for environmental science education and the National Science Education Standards. *Science Education, 91*(4), 554-578.

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*, 1–19.

Simpson, R. D., & Oliver, J. S. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education, 74*(1), 1–18.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. MIT Press.

Stake, R. E. (1995). *The art of case study research*. SAGE.

Stichter, J. P., Laffey, J., Galyen, K., & Herzog, M. (2013). iSocial: Delivering the social competence intervention for adolescents (SCI-A) in a 3D virtual learning environment for youth with high functioning autism. *Journal of Autism and Developmental Disorders, 43*(7), 1-14.

Suthers, D., & Chu, K.-H. (2012). Multi-mediated community structure in a socio-technical network. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 43–53). ACM.

Suthers, D., Fusco, J., Schank, P., Chu, K. H., & Schlager, M. (2013). Discovery of community structures in a heterogeneous professional online network. *In System Sciences (HICSS), 2013 46th Hawaii International Conference on* (pp. 3262–3271). IEEE.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. T. (2006). Planning early for careers in science. *Science, 312*, 1143–1144.

Takeuchi, L. M., & Vaala, S. (2014). Level up learning: A national survey on teaching with digital games. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Toulmin, S. E. (1958). *The uses of argument.* Cambridge University Press.

Tyson, W. (2011). Modeling engineering degree attainment using high school and college physics and calculus coursetaking and achievement. *Journal of Engineering Education, 100*(4), 760–777.

Vedder-Weiss, D., & Fortus, D. (2011). Adolescents' declining motivation to learn: Inevitable or not? *Journal of Research in Science Teaching, 48*(2), 199–216.

Walker, J. P., & Sampson, V. (2013). Learning to Argue and Arguing to Learn: Argument-Driven Inquiry as a Way to Help Undergraduate Chemistry Students Learn How to Construct Arguments and Engage in Argumentation During a Laboratory Course. *Journal of Research in Science Teaching*, *50*(5), 561–596. https://doi.org/10.1002/tea.21082

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In

Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 48-56). ACM.

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.

Xing, W., Goggins, S., & Introne, J. (2018). Quantifying the effect of informational support on membership retention in online communities through large-scale data analytics. *Computers in Human Behavior*, *86*, 227–234.

Xing, W., & Goggins, S. P. (2015). Student assessment in small groups: A spectral clustering model. *IConference 2015 Proceedings*.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168–181.

Xing, W., Wadholm, R., Petakovic, E., & Goggins, S. (2015). Group learning assessment: Developing a theory-informed analytics. *Journal of Educational Technology & Society*, *18*(2).