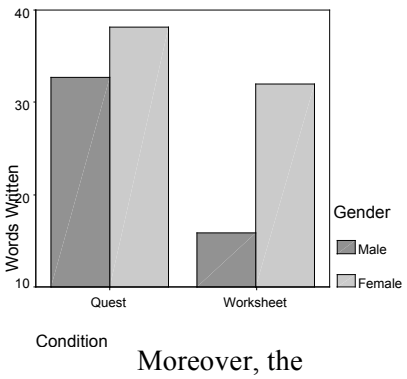
Appendix III – Sample Quest Atlantis Quantitative Studies of Student Performance – Indiana University QA Research

Learning Gains Overview

Quest Atlantis has become a highly successful game-based learning environment largely due to funding from previous funding from the National Science Foundation, the MacArthur Foundation, the National Aeronautics and Space Administration, and Food Lion. This learning and teaching project has been by used by over ten thousand 4th-8th graders worldwide on four different continents, resulting in standardized test gains, scores of research publications, innovative theoretical frameworks, the development of new media literacies, and transformative experiences that involve both game-based and real-world narratives. While we have published multiple articles and case studies discussing the impact on learners or the design principles and curriculum insights derived from particular studies, here we overview some of the quantitative learning gains that have been documented.

Language Arts / Reading Study. In a series of experiments evaluating the impact of QA on learning, we found that QA is not only engaging to children, it also solidly supports learning, and does so in authentic tasks that result in meaningful and lasting learning. For example, a Language Arts study compared student empathy toward a fictional character when the story was presented within the QA context as opposed to presenting it simply on a worksheet. Students were presented with a personal narrative and asked to respond to four open-ended questions. The Quest Atlantis condition included a 118-word introduction framing the activity with the fictional goal of aiding the Atlanteans, while the control condition presented the same first-person narrative and questions, but with only a brief, decontextualized introduction and unadorned presentation. Student responses were analyzed for content, with a .89 inter-rater reliability. While student responses to several of the questions were similar for both conditions, when asked to engage in perspective taking, *students in the QA condition offered character insights that were either deeper or better supported than did students in the worksheet condition* ($t(14) = 2.62$, $p < .05$). Considering the developmental and epistemological importance of perspective taking, the QA program may bear important benefits.



Further, in examining gender differences, girls, rather than being alienated by the videogame context, wrote significantly more ($M = 137$ words) than did boys ($M = 85$ words) ($t(18) = 2.55$, $p < .05$). Moreover, the QA context seems to have positively benefited the boys also, as they wrote more in the Quest condition than the worksheet condition (ES=.74), representing a medium-to-large effect size. Indeed, on the question concerning perspective taking, both boys and girls wrote significantly more in the QA condition (boys $M = 32.67$ words, girls $M = 38.20$) than the worksheet condition (boys $M = 15.88$ words, girls $M = 32.00$) ($t(18) = 2.17$, $p < .05$). Additionally, in terms of the question regarding perspective taking, girls (QA $M = 4.50$, Worksheet $M = 4.33$) overall offered character insights that were either deeper or better supported than boys (QA $M = 4.00$, Worksheet

$M = 3.50$), regardless of condition (t (18) = 2.55, $p < .05$). Interestingly, while boys wrote less in the worksheet condition, boys in the QA condition wrote as much as girls wrote in the worksheet condition (see Figure 2); considering that boys in general write less than girls do, *this suggests that QA can promote student achievement in a gender equitable manner.*

Social Studies Classroom Study. The Black Rhino Unit establishes a simulation context in which students investigated the socio-political and environmental dynamics surrounding the creation and maintenance of a game reserve located in the East African country of Tanzania. As part of the simulation context, children investigated and adopted multiple roles (i.e. conservationist, veterinarian, plantation owner, director of wildlife, import/export trader, meat company manager). Regardless of their particular role, children were required to develop a rich understanding of the issue at hand, a task that required them to learn about various scientific and economic aspects of the reserve and the Black Rhino, and then make recommendations about the best use of the land. As part of their participation in the project, participants traveled through the virtual space and interviewed characters who allowed them to gain access to the multi-media materials so that they could defend their positions and prepare a scientific report. The unit contained three Quests and four classroom discussion activities.

The unit as a whole took approximately six 50-minute class periods to complete. In addition to learning about the reserve in Tanzania, students also had to identify and investigate an issue in their local community. They were required to write a scientific report, prepare a persuasive argument, and develop a public document to promote awareness of this issue. We specifically examined student learning in four classrooms. In all four classrooms, there were significant learning gains from pre-test to posttest. There were significant gains from pretest to posttest at both suburban research sites (PreM = 12.14, PostM = 18.74; t(27) = 12.17, p < .01) (PreM = 6.82, PostM = 13.65; t(16) = 9.23, p < .01), at an innercity high ESL site (PreM = 2.66, PostM = 11.27; t(29) = 12.66, p < .01), and at an after-school site (PreM = 6.12, PostM = 11.38; t(7) = 10.45, p < .05). *One particular observation was that students went from almost no appreciation for how the content related to their own life (PreM = 10.30) to a deep appreciation of its relevance to their life worlds and showed the ability to take multiple perspectives in the international arena (PostM =47.45) t (19) =10.28, p<.01)* (Barab et al., in press).

As one example of the types of changes occurring between pretest to posttest, the following illustrates a sixth grader's movement from a relatively superficial description to one of much more depth.

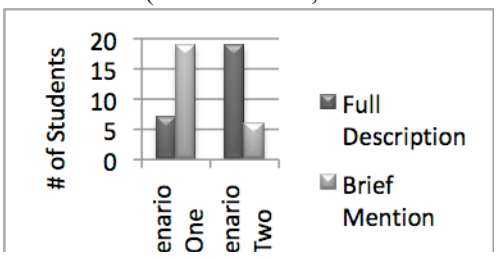
> Pre-test Response: The trade of illegal drugs is an important issue. Poor farming families know they can make money off of selling illegal substances. Drug usage is dangerous and this is an important issue.
>
> Posttest Response: In many countries, rainforest logging is a major issue. People from wealthy countries such as our own might protest it because it kills so much of our beautiful environment, but in a country where fine rainforest wood is a major industry and especially if the country's economy is weak, it's not really fair to say they can't do it anymore. This is a very controversial issue, because we are basically weighing human life and animal life, two things that depend on each other.

While the pre-test shows an attempt to respond with a factual issue, it lacks the depth and complexity of the post-test, which shows multiple perspectives, considers economic and societal issues, and weighs contradicting factors.

Language Arts / Writing Comparison Study. Within this unit, students engaged a rich narrative about a small town that is facing economic and criminal challenges such as flagging tourism, closing businesses, graffiti, and arson (Warren, Barab, & Dondlinger, in press). The unit's underlying focus is on improvement of reading and writing, as well as science learning. The student's role is one of an investigative reporter for the local newspaper. However, as soon as students engage with their first major writing task, the multi- and inter-disciplinary nature of the town's problems, solutions, and available student tasks is presented through a series of evolving, linked mysteries that students can solve. Similar to Rhino, new information and content area guidance is provided through various non-player characters (NPC) who act as pedagogical agents to present as well as help evolve the intertwining narratives that are present throughout the Anytown experience. Our analyses of Anytown have focused on learning outcomes related to writing process. Achievement scores were assessed using a standardized writing prompt used by the state of New Jersey. This prompt is open-ended and a standardized rubric is then applied to their writing responses to the prompt. For this study, a comparison, traditional curricula was made focusing on the same process writing skills, and two equivalent 5th grade classrooms were identified for this study (Warren, 2006).

We found a statistically significant increase in pre-post learning gains on the essay items for both classes, with the class using the Anytown unit showing significantly more improvement ($F(1, 40) = 4.32$, $p < .05$) than the class using the traditional curriculum. We also examined teacher time spent on administrative tasks such as providing repeated directions about the nature of the task and procedures for task completion and voluntary activities completed during the implementation. With the alpha set at .05, the paired-C-sample t-test showed that there were significant differences (t (15) =5.95, p = .043) between treatment (M = 12.12, SD = 6.70) and comparison (M = 28.41, SD = 3.91). In terms of voluntarily writing, equivalent activities were designed for both classes and students were told that they were free to engage these activities at any point. With the alpha set at .05, the paired-sample t-test showed that there are significant differences (t (40) = -16.41, p =.006 between the treatment (M = 1.09, SD = .29) and comparison class (M = .00, SD = .00).

Mathematics/Statistics. The statistics unit is one of the newer mathematics units in Quest Atlantis, and thus it has not yet been a part of a controlled comparison study. However, the unit has now been used in several classrooms, and preliminary analyses suggest that it will be a resource for powerful new understandings for students. The purpose of the statistics unit is to create opportunities for students to engage with mathematics differently; by using mathematical tools as resources for making decisions, and to recognize that mathematical decisions have real-world implications. These goals are quite different from most statistics curricula, which focus primarily on supporting students to be able to calculate statistical measures (such as mean, median, or mode) accurately. In contrast, this unit is designed to interrogate the *usefulness* of these measures in solving problems and making recommendations. *Analyses*



from preliminary implementations suggest that participating students were developing new ways of conducting analyses, and in particular, were becoming more facile with the use of justification.

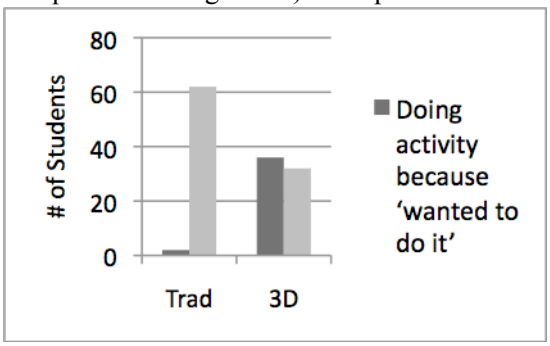
Importantly, students were consistently linking the decisions that they made based on their mathematical analyses with the real-world implications of those decisions. For example, students moved from making claims such as: "the mean of Brand A was smaller," to claims such as: "the mean of Brand A is smaller, which means that it had a shorted average braking distance. If I am worried about a safe bike, I want one that will stop quickly, so I would recommend going with Brand A." Looking across the first and last scenarios, we see significantly stronger connections of their recommendation to the scenario from scenario one to scenario two. In this way, students' were able to link their mathematical activity with contexts-of-use, and began to see mathematical procedures as tools for decision-making, rather than rules to be mastered. As one student claimed in a post-implementation interview: "I learned what math is really used for, in the real world. Before this, I didn't know when I would ever use this stuff. Now I know that things like mean, or median, can really help me out in the real world." These data are especially significant when juxtaposed with the increasing amounts of evidence indicating that students are failing to connect mathematical procedures demonstrated on school tests to real-world scenarios.

Science Classroom Comparison Study. In this research study, we sought to examine the impact and differences in teaching a science water quality curriculum in two very different contexts; one was a traditional science classroom context, and the other was the same curriculum embedded in the QA narrative. We designed a virtual watershed habitat, Taiga Park, to teach students about water quality by embedding them in a rich socio-scientific problem scenario. The Traditional condition curriculum was text-based, and activities were teacher-led. The 3D MUVE condition was virtual-based, and placed the teacher in the role of resource, while students uncovered information within the larger structure of the Quest Atlantis virtual world. Both conditions were taught by the same teacher, presented with the same domain content, and assignments were aligned so that they were highly similar in their tasks. However, the traditional condition's lessons involved lecture, class discussion, and nicely organized notes, with micro-contexts serving as examples (a new exemplar for each new concept taught, as typically found in textbooks). The 3D MUVE condition's lessons were embedded within a macro-context, the story of Taiga, where all science content was distributed in the game, needing to be uncovered and solved, within the overarching narrative about the fish decline in a local river and which groups might be responsible.

The pretest of standardized science test items showed no significant differences between conditions [$t(91) = .16$, $p = .87$, nor between classes within the conditions [$F(3, 89) = 1.09$, $p = .36$], indicating equivalent ability with respect to the target content. The post test showed significant learning gains for both conditions, indicating that both groups learned science domain concepts in this unit. However, the delayed posttest, administered at an 8 week delay, showed the 3D MUVE groups ($M = 23.65$, $SD = 5.85$) scoring significantly higher than the traditional groups ($M = 18.4$, $SD = 6.52$) both when collapsing across group membership [$t(91) = 4.02$, $p < .001$] and between classes within the conditions [$F(3, 87) = 5.73$, $p < .001$]. These results show significantly less forgetting in the 3D MUVE group, maintained over 2 months following the

intervention**. *Stated differently, the 3D game group was able to recall significantly more science concepts than the traditional students could over time*.**

Twice during the study, students were interrupted from their activities (both groups doing an equivalent assignment) to respond to a series of questions on their current state of engagement in the task at hand. This Engagement Questionnaire was based on that of Csikszentmihalyi's (1990) study with 'flow', where he interrupted students involved in various activities to respond to their current state of engagement, motivation and challenge in the task. From this scale, a variable of engagement was created by collapsing the questions of "Is this activity exciting?", "Do you enjoy what you are doing?", and "Do you wish you were doing something else?" The Internal consistency estimate for these three items was $\alpha = .88$. Comparison between conditions showed that the 3D MUVE groups ($M = 6.50$, $SD = 2.10$) scored significantly higher engagement and enjoyment than the traditional groups ($M = 2.64$, $SD = 1.94$) when collapsing across group membership [$t(104) = 9.73$, $p < .001$]. Also of interest, when we loaded both the learning gain scores and engagement reports in a discriminant function analysis, they were able to explain 83% of the variance with each variable explaining a significant amount (43% and 40% respectively) of the variance in group membership.



When asked "What is your main reason for doing this task?", 36% of the QA students chose the category 'being interested in the task' or other (18%), while the majority of traditional students (65%) reported doing their similar task 'to get a good grade' and another 30% did it because "the teacher told me to," compared to only 12% of QA students. These numbers differed significantly from chance, Chi Square = 30.38, $p < .01$. Similarly, when asked why you were doing this activity, 98% of the traditional said because they were required while only 54% of the 3D MUVE students selected this reason with 46% selecting that they were doing the activity because "they wanted to be doing it," not because required. Again, this was significantly different, Chi Square =24.87, $p < .001$. *So, in summary, we found that students who utilized the 3D MUVE were more motivated by a developing interest in the curricular content and an enjoyment of the learning experience, rather than by a sense of obligation or concern for a particular grade.*

Additional measures of engagement included the degree to which students opted to participate in non-required activities, framed as optional quests in the 3D condition or extra credit assignments in the traditional condition. These tasks were highly similar in nature, however, the rate of completion varied greatly between the two groups. In the 3D group, 38 out of 51 (74.5%) opted to do the optional quests, while in the traditional group only 2 out of 54 (3.7 %)) did the 'extra credit'. Moreover, all but 5 students in the 3D group logged on to QA after school had ended (91%), many completed additional unassigned learning Quests from home (16 %), and 14% went so far as to request an after school 'job' within the QA community.

References

Arici, A. (2008, June). Gaming the Classroom: A comparison of Learning and Engagement in a 3D Multi-User Virtual Environment and a Traditional Classroom Environment. Paper to be presented at the *International Conference of the Learning Sciences*, Netherlands.

Barab, S. A., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E-J., Kouper, I., & Herring, S. C. (2007). Situationally Embodied Curriculum: Relating Formalisms and Contexts. *Science Education*, *91*(5), 750-782.

Barab, S, A., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology*, *16*(1), 1-3.

Barab, S. A., Thomas, M, Dodge, T., Carteaux, R., and Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development 53*(1), 86-108.

Barab, S. A., Arici, A., Jackson, C. (2005). Eat your vegetables and do your homework: A design-based investigation of enjoyment and meaning in learning. *Educational Technology* 65(1), 15-21.

Barab, S., Dodge, T., Tuzun, H., Job-Sluder, K., Jackson, C., Arici, A., Job-Sluder, L., Carteaux, R., Jr., Gilbertson, J., & Heiselt, C. (in press). The Quest Atlantis Project: A socially-responsive play space for learning. In B. E. Shelton & D. Wiley (Eds.), *The Educational Design and Use of Simulation Computer Games*. Rotterdam, Netherlands: Sense Publishers.

Gresalfi, M.S. & Ingram-Goble, A. (2008). Using statistics to support arguments: The development of mathematical dispositions. Paper presented at the *American Educational Research Association Annual Conference*, New York, NY.

Sadler, T., D., Barab, S. A., & Scott, B. (in press). What do students gain by engaging in socioscientific inquiry? To appear in *Research in Science Education*.

Warren, S., Barab, S. A., & Dondlinger, M. J. (in press). A MUVE Towards PBL Writing: Effects of a digital learning environment designed to improve elementary student writing. To appear in the *Journal of Research on Technology in Education*.