It’s Better to Talk With Honey Than Vinegar: Insights Into Collaborative Learning Within Mobile AR Games

Denise M. Bressler, Greater New York City area

Introduction

According to the National Research Council [(NRC), 2012a], the ability to collaboratively solve problems is of the utmost importance in scientific careers. According to K-12 science framework authored by the NRC (2012a), “science is fundamentally a social enterprise, and scientific knowledge advances through collaboration and in the context of a social system with well-developed norms” (p. 27). The job prospects in science and technology are growing (Lockard & Wolf, 2012). However, our students are underprepared for the job requirements because these collaborative scientific practices are not cultivated in the majority of U.S. schools; when we do not prepare our students adequately for the workplace, then our national prosperity suffers (NRC, 2012b). To make the US globally competitive in science and technology, students need to be engaged with science education, build a suite of scientific practices, and learn to collaborate successfully.

Research on collaborative educational games has shown that gameplay positively impacts the development of collaboration skills (Sánchez & Olivares, 2011) and player’s perceptions of their social interactions (Mansour & El-Said, 2009). Specifically, students enjoy playing collaboratively because it encourages discussion amongst players (Sharritt, 2008). The sociocultural learning that takes place within the game works best when there is shared power and authority through scripted collaboration (Demetriadis, Tsiatsos, & Karakostas, 2012).

Within mobile learning science games, researchers have found that interdependent roles are an effective way to scaffold collaborative problem solving (Dunleavy, Dede, & Mitchell, 2009; Squire & Jan, 2007). By incorporating such interdependency, collaborative mobile augmented reality (AR) games rely on the social interactions amongst players as a key to the overall success of the games. As summarized by Klopfer (2008), students playing collaborative mobile learning games “help each other, observe each other, and act together to create communities as they learn to solve problems” (p. 223). Overall, research indicates that collaborative mobile games hold promise for promoting effective collaborative scientific practice by scaffolding and supporting discourse during gameplay.

This study investigated not only the scientific practices and collaborative responses of those playing a mobile AR game but also of those participating in a similar non-game-based activity. Specifically, this study assessed the collaborative practice and discourse of student teams during both the experimental game activity and the control lab activity. These questions guided the investigation:

1. How do communication responses within game teams compare to those within control teams?
2. How do scientific practices of game teams compare to those of control teams?
3. How else are treatment groups different when conversations are analyzed at the team level?

Methodology

Since the research questions stem from understanding the differences in the social process of learning within teams from different treatment groups, case study research was chosen as the analysis method (Yin, 2014). Specifically, a descriptive multiple case study approach was chosen with student teams as the unit of analysis (Miles & Huberman, 1994). Audio transcripts, photographic evidence, student reports, and field notes were compiled for within-case and cross-case analysis.

Participants were eighth grade science students from a middle school in Pennsylvania, USA. The school was located in a diverse, urban area with many low-income households. The district approved both the game and control activity as accepted curricula. Two teachers participated and taught several class periods including some control classes and other experimental classes. Since both conditions required collaborative groups, students were randomly assigned to teams consisting of three to four students. The process of selecting teams as case studies was purposeful random sampling (Patton, 2002). Since the school district used standardized math scores to track students into classes of above average, average, and below average math achievement, those categories were chosen to represent the continuum of achievement. In order to identify important common and contrasting patterns, teams were purposefully selected in order to achieve this continuum of achievement levels along with representation from both treatment groups (experiment and control). One team was randomly selected to satisfy each category for a total of six case studies.

The intervention started on September 23, 2013 and concluded on September 27, 2013. During the entire intervention, selected teams were audio recorded as well as documented with photographs and field notes. Onsite researchers took photographs to document student interactions on all implementation days. Field notes included observations of each period along with informal interviews with the teachers. In the control cases, two audio recording devices were placed in the center of the table and recorded audio data for each class period. In the experimental cases, recordings were conducted at the individual level; every participant on the team wore a lapel microphone attached to a small digital audio recording device placed inside a pocket. To ensure high-fidelity of the qualitative data, all collaborative discourse was transcribed to clearly delineate conversational turn-taking. Transcripts then went through two separate levels of coding. The first level was a priori based on the literature review, while the second was emergent coding based on close reading of the transcripts.

Overview of Treatment Conditions

The experiment was a mobile augmented reality game played on iPads using quick-response codes (QR codes) located throughout the school (see Figure 1). The control was a ‘tried and true’ hands-on lab experiment where students had to determine the components of a mystery powder by testing three known powders (cornstarch, baking soda, and sugar) with iodine, pH paper, vinegar, and heat. During both activities, students developed hypotheses, learned about acids and bases, and conducted basic physical and chemical tests to analyze data and determine the mystery powder.

Figure 1: Game team arriving to scan a QR code.

Control: Group Lab Activity

The control activity for this study was the mystery powder lab activity, a pre-existing curriculum unit in the district. Conducted early in the 8th grade school year over the course of three to five days, the activity exposes students to basic scientific practices and some content about acids and bases. With the teachers and the principal, the researcher selected this activity as the control for several reasons:

- Students engage in scientific practices described by the National Research Council (2012a).
- It is implemented as a collaborative scientific investigation with small groups of students.
- It has the element of mystery.
- It has already been taught for at least one school year.
- The content lends itself to game-based learning.

Experiment: Collaborative Mobile AR Game

Using the mystery powder lab as the starting point for the design, the content from the lab was transformed into a mobile AR game. As students moved throughout their school building, they encountered QR codes that they scanned to access game information. This included conversing with virtual characters and gaining evidence to keep in inventory. Players were also required to talk to real people in the building to get additional game information. Players also deciphered a code and typed in the answer manually to the decoder. The game was played in teams of three or four where each student had a unique role: social networker, techie, photographer, or pyro-technician. Based on their role, they were provided with different pieces of information as they progressed through the game. The roles were designed interdependently; thus, to solve the mystery, players had to share information and work together.

In the game narrative, someone stole money from the cafeteria cash register and left behind a mysterious white powder. The game took place as five chapters, roughly aligning to one chapter per class period. Chapter #1: Students were introduced to the incident and the main characters. They visited the cafeteria to explore the crime scene and then several more locations to discuss the incident with the three main suspects: the janitor, the secretary, and a fictional fellow student. Chapter #2: Students visited areas of the school where the suspects left evidence. At each location, they found evidence of the known powders and conducted some simple, virtual tests including vinegar, iodine, heat, and pH tests. Content knowledge and test results were all conveyed using pictures and videos during gameplay. Chapter #3: A sample of a real mystery powder was provided for testing. Facilitated by some teacher instruction and assistance, game teams conducted tests on an actual powder (see Figure 2). Instructional prompts were provided by a main character from the game.

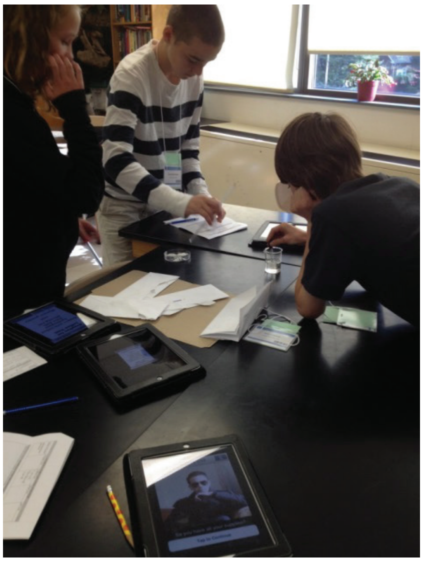

Figure 2: Mobile AR game team conducting hands-on experiment.

In Chapter #4 and Chapter #5: Teams revisited the crime scene to see if they missed anything and discovered an additional piece of data necessary to confirm the identity of the thief. Then, they revisited the locations where suspects stored their belongings and collected additional evidence. Once students determined the thief's identity, they gave their final accusation to the in-game principal.

Results

First, the within-case analysis for each team includes a brief case overview. Second, the cross-case analysis represents all cases in a meta-data matrix. The matrix is conceptually ordered: teams at the top worked together most effectively while teams near the bottom were not as effective. Finally, to answer the research questions, findings from the cross-case synthesis are discussed.

<u>Control Team #1</u>: Selected from a class with above average math achievement; this team consisted of two boys and two girls. In general, one boy did not want to do the lab, while the other one kept walking away from the group. One girl was very talkative with others, while the other was generally on-task. Over the course of the activity, no leader emerged. While their process of interaction was democratic, it was also fairly ineffective. The biggest problem for this group was their confusion.

Control Team #2: Selected from a class with average math achievement; this team consisted of two boys and two girls. First, the girls were somewhat hesitant to talk. One boy wanted to take leadership and did not want anyone else to do anything; the other boy seemed willing to defer to the leader boy. During the activity, the strong-willed boy controlled the leadership; he was a very controlling, demanding leader and an ineffective communicator. Group members disagreed often and did not support each other's ideas. Group issues seemed to stem from fighting over roles and responsibilities.

Control Team #3: Selected from a class with below average math achievement; this team consisted of two boys and one girl. In the beginning, the boys were kind of quiet. The girl seemed knowledgeable and interested in science and took a leadership role. She would delegate to the boys, yet sometimes she got aggravated with them. There was a mixed level of support for each other's ideas. The biggest problems for this group were there high level of off-topic conversations and moderate confusion.

Game Team #1: Selected from a class with above average math achievement; this team consisted of four girls. They were generally on-task and seemed to stay together and work well as a group. Over the course of the activity, no leader emerged. Instead, they discussed ideas as a group and supported each other's ideas. This team had no noticeable issues; they suffered little confusion and stayed on task towards their goal. They had the highest written report scores of any case study team.

Game Team #2: Selected from a class with average math achievement; this team consisted of three boys and one girl. In general, one boy did not seem to get along entirely well with the group. Over the course of the activity, no leader emerged. Group members disagreed about half of the time and supported each other's ideas the other half of the time. Their process of interaction was democratic and generally effective. Overall, this team struggled somewhat with group dynamics in situations that were outside of the game framework, such as conducting the lab experiment. However, when it came to synthesizing the information and drawing conclusions collectively as a group, they excelled.

Game Team #3: Selected from a class with below average math achievement; this team consisted of three boys and one girl. In general, all the individuals in this group seemed quiet and reserved; however, one boy took a leadership role and taught the rest of his group about the content and technology. The group's biggest problem may have been the reserved nature of members. The team had low conflict and low confusion; however, dynamics did not yield fully productive conversations. Overall, their process of interaction was a blend of directed leadership and communal effort.

RQ1: Communication Responses

Responses that occurred in team conversations were categorized as accept, discuss, and reject. The code structure built on the work of Barron (2003). When a student agreed with the speaker, supported the idea, or proposed a next step, the interaction was coded as *accept*. When interactions facilitated further discussion, such as questioning an idea, asking for clarification, or challenging an idea with new information, the interaction was coded as *discuss*. When a student rejected an idea or interacted in a way that would not facilitate further discussion, the interaction was coded as *reject*. Based on code reports, occurrences were categorized into levels of low (under 7), moderate-low (7-14), moderate (15-22), moderate-high (23-30), high (31-38), very high (over 38) for each response type.

When comparing communication response types between treatments, game teams and control teams showcased different patterns of communication responses (see Table 1 for occurrences). First, game teams had moderate to low levels of *reject* responses, while control teams had moderate to high levels of *reject* responses. Second, game teams had moderate to high levels of *accept* responses; while control teams had only moderate to low levels of *accept* responses. Lastly, game teams had high or very high levels of *discuss* responses; while control teams had mostly moderate levels of *discuss* responses. Barron (2003) categorized accept and discuss responses as *engaged* responses, while reject responses are considered *non-engaged* responses. Game teams produced a fairly high level of engaged responses in comparison to their non-engaged responses. In contrast, control teams produced a fairly high level of non-engaged responses in comparison to their engaged responses.

	Communication Responses	Scientific Practices	Language Style
Game Team #1 (above average math)	Discuss: High (33) Accept: High (31) Reject: Low (6)	Interpreting data: Very high (21) Constructing explanations: High (9) Arguing with evidence: Moderate (6) Defining the problem: Moderate (5) Planning investigation: Low (4)	Commands: Low (6) Communal: High (39)
Game Team #2 (average math)	Discuss: Very High (59) Accept: Mod-High (23) Reject: Moderate (18)	Interpreting data: Very high (38) Constructing explanations: High (14) Arguing with evidence: Moderate (5) Defining the problem: Moderate (6) Planning investigation: Low (4)	Commands: Moderate (23) Communal: High (63)
Game Team #3 (below average math)	Discuss: Very High (47) Accept: Mod-Low (14) Reject: Low (2)	Interpreting data: Very high (30) Defining the problem: High (11) Arguing with evidence: Moderate (5) Constructing explanations: Low (4) Planning investigation: Low (2)	Commands: Moderate (32) Communal: High (43)
Control Team #1 (above average math)	Discuss: Moderate (21) Accept: Moderate (15) Reject: Moderate (16)	Interpreting data: Very high (59) Planning investigation: High (13) Arguing with evidence: Moderate (7) Constructing explanations: Moderate (6) Defining the problem: Moderate (6)	Commands: Moderate (28) Communal: Moderate (22)
Control Team #3 (below average math)	Discuss: Mod-High (26) Accept: Moderate (21) Reject: Mod-High (26)	Interpreting data: Very high (65) Planning investigation: High (13) Constructing explanations: Low (3) Defining the problem: Low (1) Arguing with evidence: Low (2)	Commands: High (49) Communal: Low (7)
Control Team #2 (average math)	Discuss: Moderate (19) Accept: Mod-Low (13) Reject: High (37)	Interpreting data: Very high (106) Planning investigation: High (13) Constructing explanations: Moderate (7) Defining the problem: Low (3) Arguing with evidence: Low (3)	Commands: High (50) Communal: Low (17)

Table 1: Conceptually-Ordered Discourse Summary for All Cases.

RQ2: Scientific Practices

The scientific practices that occurred in team conversations were coded to align directly to the scientific practices from the National Research Council (2012a). When students discussed what was known about the investigation or tried to determine what needed to be answered, then the dialogue was coded as *Defining the Problem*. When students discussed their investigation plan or what information they needed to record, then the dialogue was coded as *Planning out the Investigation*. When students discussed characteristics of the experiments they were observing, then the dialogue was coded as *Interpreting Data*. When students tried to explain the relationships between data, then the dialogue was coded as *Constructing Explanations*. When students supported or refuted an argument by citing relevant evidence, then the dialogue was coded as *Arguing with Evidence*. Based on code reports, occurrences were categorized into levels of low (1-4), moderate (5-8), high (9-14), and very high (over 14) for each scientific practice.

When comparing scientific practices between treatments, game teams and control teams showcased different usage patterns of scientific practices during their conversations (see Table 1 for occurrences). Since reviewing the number of occurrences of each practice did not reveal the whole story, a more detailed analysis of the conversational occurrences was necessary. First, for occurrences coded as *Defining the Problem*, game teams revealed a stronger understanding of describing the problem as well as some understanding of how to create a hypothesis. While control teams did showcase this practice, they only revealed a basic understanding of describing the problem and a very basic understanding of how to create a hypothesis. Second, for occurrences coded as *Planning out the Investigation*, control teams had a better understanding of the plan they needed to execute in order to determine the identity of the mystery powder than game teams. Third, for *Interpreting Data*, although both treatments had a high level of occurrences, game teams offered observations that were more specific and substantive than control teams. Fourth, for occurrences coded as *Constructing Explanations*, the below average teams from both treatments struggled somewhat with this practice exhibiting only a basic understanding. However, in comparing the higher achieving students, game teams constructed explanations about both the game narrative and the scientific content leading to more opportunities to showcase this practice whereas control teams only explained the science content. Finally, when *Arguing with Evidence*, game teams revealed their ability to argue with evidence more than once during the activity; multiple team members were also involved in making evidence-based arguments. Not all control teams showcased this practice on their own; for those that did, they only revealed it once at the activity's end and only one control team had multiple members exhibiting the practice. Other than when *Planning out the Investigation*, conversations amongst game teams revealed a greater ability to engage in scientific practices than control teams.

RQ3: Other Differences

As mentioned earlier, some codes emerged during a second round of emergent coding. When reviewing transcripts, the researcher noticed differences in the general language style of the treatment groups. Students in the control were frequently telling each other what to do. They were using language such as, "don't reach across the table like that—here—give it" (Control Team #1, 510), "put the whole entire thing in" (Control Team #2, 122), and "go get the other one" (Control Team #3, 285). To capture this type of directive language, a new code was created called *commands*. Additionally, the researcher noticed that students in the experiment were addressing the group collectively, rather than one specific team member. They were also referring to the group as an entity with words such as "we," "we're," and "let's." To capture this type of communal language, a new code was created called *communal*. Based on code reports, occurrences were categorized into levels of low (19 and under), moderate (20-38), and high (over 38) for each language style.

In terms of language style, game teams and control teams demonstrated an emphasis on different styles during their conversations (see Table 1 for occurrences). Game teams had high levels of communal language and moderate to low levels of command language. In contrast, control teams had moderate to low levels of communal language and moderate to high levels of command language. For the entire activity, conversations amongst game teams had not only higher levels of engaged responses but also higher levels of communal language and a greater ability to engage in scientific practices. These patterns of group communication seemed to connect with better group dynamics and more effective team communication. In contrast, over the whole activity, conversations amongst control teams had not only higher levels of rejecting responses but also higher levels of commands and a reduced ability to engage in scientific practices. These patterns of group communication seemed to connect with less effective group dynamics and poor team communication skills.

Discussion and Conclusion

Prior research indicated that collaborative games held promise for promoting effective collaborative practice by scaffolding and supporting discourse during gameplay. Specifically, when it comes to scientific practice, research has showed that students guided to socially construct their knowledge in *River City* had a stronger understanding of scientific inquiry than other students (Ketelhut, Nelson, Clarke, & Dede, 2010). Similarly in this study, game teams communicated well and showcased greater levels of scientific practice. The game in this study utilized interdependent roles and jigsaw pedagogy to scaffold player's social interactions. Based on the player's role, unique information was revealed to the player that he or she had to share with others. Aronson and Patnoe (2011), experts on using jigsaw pedagogy in the classroom, argued that this style of social interdependence is a way to promote effective group learning because as members start to learn from each other, the feeling that they need to outperform their classmates diminishes. Unfortunately, control teams struggled to understand their individual roles within the group and their group dynamics suffered. They showcased ineffective communication responses and language styles, possibly due in part to the desire to outperform teammates, which resulted in lower levels of scientific practice.

According to Reiser, Berland, and Kenyon (2012), students need to "actively listen and respond to one another" in order to be engaged in meaningful scientific practice (p.36). Game teams had more engaged communications responses along with higher levels of communal language; in other words, they spoke to each other with 'honey'. In brief, game teams met the precursor for meaningful learning by communicating with honey. Unfortunately, control teams had higher rejecting responses and higher commands, thus they spoke to each other with 'vinegar'. Control teams did not meet the precursor for meaningful learning since they communicated with vinegar.

As mentioned in the beginning of this paper, the ability to collaboratively solve problems is of the utmost importance in scientific careers, yet students in most U.S. schools are not exposed to activities that promote scientific practice in an effective and engaging collaborative setting. Collaborative mobile AR games designed with interdependent roles hold promise for offering exactly this type of learning experience. The game in this study was implemented within the practical parameters of a real school setting. For schools that have iPads, this type of game could be scaled up and implemented as support for the Next Generation Science Standards (NGSS). With the recent release and adoption of the NGSS, schools will need curriculum activities that support student learning aligned to these standards. The coded dialogue for scientific practice was perfectly aligned to the NGSS and the study shows that the game teams had greater levels of scientific practices in their conversations than control teams. Game teams also showcased that scientific knowledge can be advanced through student collaboration by talking with 'honey' and not 'vinegar'. All in all, collaborative mobile AR games that are designed to promote not only NGSS but also communication skills should be strongly considered by school policy makers.

References

Aronson, E., & Patnoe, S. (2011). *Cooperation in the classroom: The jigsaw method* (3rd ed.). London: Pinter & Martin Ltd.

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences, 12*(3), 307-359.

Demetriadis, S., Tsiatsos, T., & Karakostas, A. (2012). Scripted collaboration to guide the pedagogy and architecture of digital learning games. *Proceedings of the European Conference on Games Based Learning*, 148-154.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18*(1), 7–22.

Ketelhut, D. J., Nelson, B. C., Clarke, J., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology, 41*(1), 56-68.

Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games*. Cambridge, MA: The MIT Press.

Lockard, C. B., & Wolf, M. (2012). *Occupational employment projections to 2020*. Washington, DC: U.S. Bureau of Labor Statistics. Retrieved from http://www.bls.gov/opub/mlr/2012/01/art5full.pdf

Mansour, S. S., & El-Said, M. (2009). Multi-players role-playing educational serious games: A link between fun and learning. *International Journal of Learning, 15*(11), 229-239.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

National Research Council. (2012a). *A framework for K-12 science education: Practices, cross-cutting concepts, and core ideas*. Washington, DC: The National Academies Press.

National Research Council. (2012b). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Reiser, B. J., Berland, L. K., & Kenyon, L. (2012). Engaging students in the scientific practices of explanation and argumentation. *Science Teacher, 79*(4), 34-39.

Sánchez, J., & Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education, 57*(3), 1943-1952.

Sharritt, M. J. (2008). Forms of learning in collaborative video game play. *Research & Practice in Technology Enhanced Learning, 3*(2), 97-138.

Squire, K., & Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education & Technology, 16*(1), 5-29.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage Publications, Inc.