Better Than Business-as-Usual: Improving Scientific Practices During Discourse and Writing by Playing a Collaborative Mystery Game

Denise Bressler
Learning Edge Ventures

Introduction

In 1910, John Dewey published *How We Think* and discussed his views on training students to think well. To Dewey (1910/1997), students should be taught to think independently—to reason, to interpret facts, and to think abstractly; instead, he argued, students were being trained to produce a teacher-approved answer. Today, approximately 100 years later, concern over students' lack of twenty-first century learning skills (National Research Council [NRC], 2007) reflect the same apprehensions of Dewey. The Framework for 21st Century Learning prepared by the Partnership for 21st Century Skills (2014) details several aspects of critical thinking; students should be able to analyze evidence, reason effectively, solve problems, and make connections between information.

In an effort to incorporate such critical thinking skills into science education, the NRC (2012) commissioned the K-12 science framework. The report outlined eight scientific practices that are integral to today's science education curricula. These practices include interpreting data, constructing explanations, and arguing with evidence. This list of scientific practices aligns closely to several aspects of critical thinking as detailed in The Framework for 21st Century Learning (Partnership for 21st Century Skills, 2014). Critical thinking, an important concept to Dewey, is still poorly taught and poorly assessed, particularly in science education.

One way to gauge scientific practice and critical thinking is through writing. Wallace, Hand, and Prain (2004) argued that writing fosters science learning. When students are scaffolded through the process of science writing, their ideas evolve from vague notions into specific, complex understandings (Keys, Hand, Prain, & Collins, 1999). Additionally, writing to learn is particularly powerful when combined with collaborative peer discussion (Chen, Hand, & McDowell, 2013; Keys et al., 1999).

Collaborative peer discussion is not only important to the science writing process, it is crucial for improving science education. According to the NRC (2012), the ability to collaboratively solve problems is of the utmost importance in scientific careers. The K-12 science framework authored by the NRC (2012) states that "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).

Research has shown that collaborative learning games are effective at supporting collaboration and collaborative peer discourse. Gameplay positively impacts the development of collaboration skills (Sánchez & Olivares, 2011) and players' perceptions of their social interactions (Mansour & El-Said, 2009). Specifically, students enjoy playing collaboratively because it encourages discussion among 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).

There is a burgeoning body of research on collaborative mobile augmented reality (AR) games specifically for science learning that holds promise for promoting not only collaboration but also scientific practice and critical thinking. 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 AR games rely on the social interactions among players as a key to the overall success

DOI: 10.37514/DBH-J.2014.2.1.04

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.

Collaborative mobile AR games, still in infancy, also show potential for promoting science learning (Squire & Klopfer, 2007) and scientific literacy (Squire & Jan, 2007). Mobile games designed to put students in professional roles have shown that during gameplay students engage in the process of scientific inquiry (Dunleavy et al., 2009; Rosenbaum, Klopfer, & Perry, 2007; Squire & Klopfer, 2007) and argumentation (Mathews, Holden, Jan, & Martin, 2008; Squire & Jan, 2007). Squire and Jan (2007) stated that these "games could have promise for tools that develop scientific literacy" (p. 23). Specifically, collaborative mobile games hold promise for promoting critical thinking as embodied by such scientific practices.

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. The activities were designed to support science writing in the form of open-ended explanations on an incident report. The game was designed to support collaborative discourse through the use of interdependent roles, while the non-game activity simply put students in groups. Specifically, this study assessed the science writing and collaborative 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 of game teams compare to those of control teams?
2. How do scientific practices of game teams compare to those of control teams?
3. How else are treatment groups different when discourse is 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 qualitative method of analysis (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, U.S.A. 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 3 to 4 students.

The process of team selection for the qualitative strand 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 selected in order to achieve a continuum of achievement levels (above average, average, and below average) and representation from both treatment groups (experiment and control), as shown in Table 1. One team was randomly selected to satisfy each classification in order to increase the credibility of the findings.

Table 1 *Purposeful Random Selection of Student Teams*

	Above Average	Average	Below Average
Control	Team C1	Team C2	Team C3
Experiment	Team E1	Team E2	Team E3

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. For the a priori coding, code sets were used to investigate (1) communication responses and (2) scientific practices. First, Barron (2003) found that successful collaborative teams offer significantly more engaged responses than less successful teams; therefore, the first set of codes replicated her code structure of (a) accept, (b) discuss, and (c) reject. Second, Squire and Jan (2007) determined that mobile AR games can promote scientific argumentation; their codes included question, hypothesis, counter-hypothesis, and evidence. This research study expanded on their code structure by utilizing codes that align directly with the scientific practices published by the NRC (2012). For the emergent coding, the researcher created memos with ideas for new themes. Then, new codes were developed to further interpret team interactions. In order to explain the patterns of engaged responses among treatment groups, codes were created for *commands* and *communal language*.

Written reports requiring open-ended explanations were coded independently by two individual graders using a rubric. The rubric enabled grading of written statements on a Likert-style scale where "insufficient response" was equal to 0 and "exemplary response" was equal to 4. Students were graded on six of the eight scientific practices as outlined by the NRC (2012). The six practices were

- Practice #1: Asking questions and defining problems
- Practice #3: Planning and carrying out investigations
- Practice #4: Analyzing and interpreting data
- Practice #6: Constructing explanations
- Practice #7: Engaging in argument from evidence
- Practice #8: Obtaining, evaluating, and communicating information.

The instrument contained individual items pertaining to each practice. Students were rated on a total of 9 items; possible score range was 0 to 36. The two coders met for several sessions to discuss all mismatches until unanimous consensus was eventually achieved.

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 in which 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 (see Figure 2). Conducted early in the eighth-grade school year over the course of three to five days, the mystery powder lab activity exposes students to basic scientific practices and promotes a foundational understanding of acids and bases. In concert 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 (2012).
- It is implemented as a collaborative scientific investigation with small groups of students.
- It has the element of mystery which is an important parallel to the game's narrative.
- 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, 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 even deciphered a code and typed in the answer manually to the decoder.


Figure 2. Student teams participating in mystery powder lab (control condition).

The game was played in teams of three or four in which each student had a unique role: social networker, techie, photographer, or pyro-technician. Based on their respective roles, they were provided with different pieces of information as they progressed through the game. The roles were designed interdependently; therefore, 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 where they found the cash register broken into— and a mystery powder left behind! Then, several more locations were visited to interview the main suspects: the janitor, the secretary, and a fictional fellow student. Chapter #2: Students visited areas of the school where suspects left evidence. At each location, they found evidence of the known powders (cornstarch, baking soda, and sugar) and conducted some simple, virtual tests including vinegar, iodine, heat, and pH tests. Powders and tests were exactly the same as for the control group. The difference was that content knowledge and tests results were all conveyed using pictures and videos during gameplay. Chapter #3: The chief detective, a character in the game, finally cleared the mystery powder for testing. A sample of a real mystery powder was provided at this time and the detective described how to run each test. Facilitated by some teacher instruction and assistance, game teams conducted tests on an actual powder (see Figure 3).

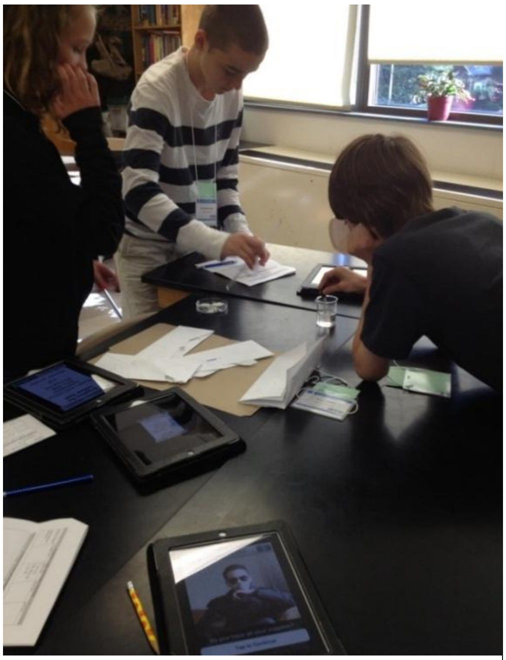

Figure 3. Game team conducting hands-on experiment.

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 made 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.

Team #C1: This team was a randomly selected control team with above average math achievement; it consisted of two boys and two girls. In general, one boy did not want to do anything, while the other one kept disappearing and walking away from the group. One of the girls was very talkative with other people, while one girl 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. All reports showed a lack of understanding about describing and planning the experiment; all students wrote their own answers and displayed some proficiency in data interpretation and constructing explanations.

Team #C2: This team was a randomly selected control team with average math achievement; it consisted of two boys and two girls. In the beginning, 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. Over the course of 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. All reports were almost identical; during the activity, the students discussed copying answers several times. Reports were weak on asking questions and planning out the experiment; however, they were proficient at constructing explanations and exemplary on data interpretation.

Team #C3: This team was a randomly selected control team with below average math achievement; it consisted of two boys and one girl. In the beginning, the boys were fairly 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 their high level of off-topic conversations and moderate confusion. All reports showed a lack of understanding about describing and planning the experiment as well as writing hypotheses; reports were between developing and proficient for data interpretation and constructing explanations. Data tables showed evidence of mismanaging observations, as evidenced in their dialogue.

Team #E1: This team was a randomly selected game team with above average math achievement; it consisted of four girls. The girls were generally on-task and seemed to stay together and work well together 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. All reports were scored as exemplary for describing the incident and data interpretation; they were scored as proficient for planning out the investigation. The girls had a developing capacity for writing hypotheses and were proficient at constructing explanations about connections between data points.

Team #E2: This team was a randomly selected game team with average math achievement; it 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, they excelled at synthesizing the information and drawing conclusions collectively as a group. All reports were exemplary for describing the incident and data interpretation along with a developing capacity for writing hypotheses. This team was weak on planning out their investigation. For the scientific practice of constructing explanations, reports showed a range from beginning to exemplary; their ability to write about connections within the experimental results varied greatly.

Team #E3: This team was a randomly selected game team with below average math achievement; it 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 technology and the content. The group's biggest problem may have been the reserved nature of the members. The team had low conflict and low confusion; however, the dynamics did not yield the most productive conversations. Overall, their process of interaction was a blend of directed leadership and communal effort. All reports were exemplary for data interpretation and showed a developing capacity for describing the incident and writing hypotheses. Similar to Team #E2, reports showed a range from beginning to exemplary for the practice of constructing explanations; this team revealed varying abilities to write about connections between data. Reports were weak on planning out their investigation; students did not think through the necessary steps.

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). For examples of student dialogue coded with accept, discuss, and reject, please refer to Table 2. 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 5 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.

Table 2 *Coding Definitions for Communication Responses with Examples*

Response	Definition	Examples
Accept	When a student agreed with the speaker, supported the idea, or proposed a next step	S1: He said there's no wrong answer. S2: Exactly. There's no wrong answer. (*Team C1 Conversation, 109:110*) S1: Okay, wait. Do we dip this in? S2: Yeah, when it's like liquid. (*Team C2 Conversation, 418:419*) S1: Number 7? S2: Yeah. (*Team C3 Conversation, 1109:1110*).
Discuss	When a student questioned an idea, asked for clarification, or challenged an idea with new information	S1: Yeah, it turned black and hardened. S2: So maybe cornstarch and sugar. (*Team C1 Conversation, 944:945*) S1: Why aren't we doing B? S2: Because he picked A. We can just do whichever one we want, but William picked A. (*Team C2 Conversation, 1589:1590*) S1: Where do we write that? S2: Middle of the box. (*Team C3 Conversation, 10:11*)
Reject	When a student rejected an idea or interacted in any way that would not facilitate further discussion	S1: It's fine. S2: No, it isn't. (*Team C1 Conversation, 183:185*) S1: Wait, can I see it real quick? S2: No, it's my turn. It's my turn. (*Team C2 Conversation, 903:905*) S1: It's shrinking. S2: It can't shrink unless it's melting. (*Team C3 Conversation, 627:628*)

RQ2: Scientific Practices

Discourse in team conversations and open-ended explanations on written reports were coded to align directly with the scientific practices from the National Research Council (2012). For each practice, the type of dialogue that would qualify as representing that practice was defined. For coding definitions of each practice along with examples of student dialogue, refer to Table 3. 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 5 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 revealed only 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

Table 3 *Coding Definitions for Scientific Practices with Examples*

Name	Definition	Example
Defining the Problem	When students tried to determine what needed to be answered or discussed what was known about the investigation	We will test three powders, do various experiments to ultimately conclude which powder is which. (*Team C1 Conversation, 41*) There was a theft at the cafeteria. (*Team E2 Conversation, 45*)
Planning the Investigation	When students discussed their investigation plan or what information they needed to record	Okay, using the first powder we can get, we will do the four tests - heat, pH, vinegar and iodine - and then keeping them in that order. (*Team C2 Conversation, 32*) Write the steps that you'll perform to identify the mystery powder. Well let's just say go talk to the janitor. (*Team E1 Conversation, 89*)
Interpreting Data	When students discussed characteristics of the experiments they were observing	It looks like boogers and snot. (*Team C2 Conversation, 405*) It's turning black. (*Team C3 Conversation, 352*) This is the baking soda one. So for that one it reacted and started to bubble. (*Team E3 Conversation, 802*) It turns yellow and then it turns liquidy and white again. (*Team E2 Conversation, 723*)
Constructing Explanations	When students constructed an explanation in order to explain the relationship between data	It was just bubbling like the baking soda. (*Team C2 Conversation, 1149*) Okay, well, this is fizzing. The only one that fizzed is baking soda. (*Team E1 Conversation, 581:582*)
Arguing with Evidence	When students supported or refuted an argument by citing relevant evidence	I think it's going to be sugar and baking soda, because it smelled…and it was just bubbling like the baking soda. (*Team C2 Conversation, 1149*) The cornstarch burned. So the only thing that burns is the cornstarch, so it has to be the cornstarch. (*Team E2 Conversation, 1351*)

substantive than control teams. Fourth, for occurrences coded as *Constructing Explanations*, the below-average achievement teams from both treatments struggled somewhat with this practice, exhibiting only a basic understanding. However, higher achieving game teams constructed explanations about both the game narrative and the scientific content, leading to more opportunities to showcase this practice, whereas higher achieving 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 revealed it only 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 among game teams revealed a greater ability to engage in scientific practices than control teams.

Open-ended explanations in written reports were also coded to align directly to the scientific practices from the National Research Council (2012). For illustrative examples from both treatment groups, refer to Table 4. First, most writing generated by game teams *defined the problem* in an exemplary capacity; this is similar to occurrences of this scientific practice found in game team discourse. Second, in their writing, game teams exhibited a developing understanding of how to *write a hypothesis*; this is also similar to how this practice occurred during discourse. Third, game teams all showed an exemplary ability to *interpret data* in their written reports; observational descriptions were detailed and specific. Similarly, discourse about data within game teams was more explicit and substantive than control teams; it clearly impacted their ability to write effectively about the data. Fourth, student reports from both the control and the game teams revealed a low-level capacity for *planning out the investigation*; students need more practice with this skill. Finally, there was a range of capacities exhibited for *constructing arguments* in writing; it appears to be a skill that is highly individualized to each student. More data is needed to draw conclusions about the connection between team discourse and students' ability to construct explanations and draw connections between data points in their writing.

RQ3: Other Differences

As mentioned earlier, some codes emerged during a second round of 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 5 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 among 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 among 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.

Table 4 *Science Writing with Illustrative Examples*

Practice	Team	Example
Defining the Problem	Game	The money from the cafeteria register was stolen and the only evidence is a mystery powder. The janitor is a suspect. He needed money for his tumor procedure and he uses white powder to clean. Destiny also because she needed money for the trip and her rocket uses the same powder.
	Control	What two powders are in the box. You have to test things in order to find out what the powders are.
Planning the Investigation	Game	If the mystery powder fizzes and another powder that we check fizzes than that is the mystery powder.
	Control	Using the first powder we can get, we will do the heat, pH, vinegar, and iodine in that order. And so on…
Write a hypothesis	Game	If I add vinegar, the mystery powder will fizz If I add iodine, the mystery powder will turn orange and slushy If I add heat, the mystery powder won’t do anything If I test pH, the mystery powder will read as a 8
	Control	If I add vinegar, the mystery powder might bubble up If I add iodine, the mystery powder will turn a color If I add heat, the mystery powder will maybe smoke or burn If I test pH, the mystery powder will match a color
Interpreting Data	Game	Vinegar: started to fizz Iodine: clumps to iodine, blackens, hardens, turns to clumps Heat: burned, turned black and brown pH: 8
	Control	Vinegar: bubbled up Iodine: turned black and dried up Heat: burned, turned black, and made smoke (fog) pH: turned green
Constructing Explanations	Game	With the vinegar test, it fizzed. With the iodine test, it turned hard and black. With the heat test is rised and burnt. With the pH test is was pH 8. The mystery powder is the cornstarch because it matches the mystery powder the best.
	Control	The vinegar test. The powder started to fizz and that was a sign of baking soda. Iodine test. It turned black and that was a sign of cornstarch. So it is cornstarch and baking soda.

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 in both their writing and discourse. The game in this study utilized interdependent roles and jigsaw pedagogy to scaffold the players' 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

Table 5 *Conceptually-Ordered Discourse Summary for All Cases*

	Communication Responses	Scientific Practices	Language Style
Team #E1	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)
Team #E2	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)
Team #E3	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)
Team #C1	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)
Team #C3	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)
Team #C2	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)

Note. Occurrences appear in parentheses. Scale to determine communication level was low (6 and under), moderate-low (7-14), moderate (15-22), moderate-high (23-30), high (31-38), very high (over 39). Scale to determine scientific practice was low (1-4), moderate (5-8), high (9-14), very high (over 14). Scale to determine levels of commands and communal language was low (19 and under), moderate (20-38), high (over 38).

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 in both their writing and discourse.

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 brief, game teams met the precursor for meaningful learning by communicating effectively. Unfortunately, control teams had higher rejecting responses and higher commands. Control teams did not meet the precursor for meaningful learning since they communicated poorly.

Further proof of the results of such effective and ineffective collaborative discourse was revealed in the written reports. Gamers had a stronger capacity to describe the problem as well as analyze and interpret data; in general, their reports yielded higher scores and they did not have to copy answers in order to achieve such scores. Their conversational dialogue showcased effective scientific practices and critical thinking, and their reports captured this socially constructed learning.

As mentioned earlier, critical thinking is poorly taught and poorly assessed, particularly in science education. Collaborative mobile AR games designed with interdependent roles hold promise for offering a learning experience that promotes critical thinking, as seen both in students' science writing and collaborative discourse. 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. NGSS are the curriculum standards that derived from the K-12 Framework commissioned by the NRC (2012). In this study, students' science writing was coded for scientific practices that perfectly aligned to the NGSS, and the study showed that the game teams had greater levels of scientific practices in their conversations than control teams. Game teams also demonstrated that scientific knowledge can be advanced through student collaboration by talking with communal language rather than command language. 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.

Chen, Y.-C., Hand, B., & McDowell, L. (2013). The effects of writing-to-learn activities on elementary students' conceptual understanding: Learning about force and motion through writing to older peers. *Science Education, 97*(5), 745 – 771.

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.

Dewey, J. (1997). *How we think*. Mineola, NY: Dover Publications. (Original work published 1910).

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.

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary science. *Journal of Research in Science Teaching*, *36*(10), 1065 – 1084.

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

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.

Mathews, J., Holden, C., Jan, M.-F., & Martin, J. (2008). *Sick at South Shore Beach: A place-based augmented reality game as a framework for building evidence-based arguments*. Paper presented at the Proceedings of the 8th international conference on International conference for the learning sciences - Volume 3, Utrecht, The Netherlands.

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

National Research Council. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: The National Academies Press.

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

Partnership for 21st Century Skills. (2014). *Framework for 21st Century Learning* [Web page]. Retrieved from http://www.p21.org/our-work/p21-framework

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.

Rosenbaum, E., Klopfer, E., & Perry, J. (2007). On location learning: Authentic applied science with networked augmented realities. *Journal of Science Education & Technology, 16*(1), 31 – 45.

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.

Squire, K., & Klopfer, E. (2007). Augmented reality simulations on handheld computers. *Journal of the Learning Sciences, 16*(3), 371 – 413.

Wallace, C., Hand, B., & Prain, V. (2004). *Writing and learning in the science classroom*. Dordrecht, Holland: Kluwer.

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