Abstract: This paper presents two cycles of a design-based research project that considers whether and how consequential feedback supports students’ engagement with mathematics. Consequential feedback describes a form of feedback that is embedded in the context with which a student is engaging, and allows the student to see how their solution to a problem plays out in that context. Comparing two classes across two years, we considered how the timing of feedback impacted student engagement. Findings suggest that providing consequential feedback in the form of a narrative outcome supported students to offer more mathematical justification, consequential justification, and to engage critically with the mathematical content within their written recommendations. The impact of consequential feedback appears to be potentially heightened by total time in class discussion, presence of more frequent discussions, and increased immersion in the narrative.

Introduction and Framing

In this paper, we present findings from the first two rounds of a design-based experiment that sought to consider how particular forms of feedback might effectively support productive forms of mathematical engagement. The conjecture underlying this research is that a particular form of feedback, called consequential feedback, can serve to both help students to interrogate their own mathematical work and provide information for students to use to improve their performance. Consequential feedback describes a form of feedback that is embedded in the context (mathematical or narrative) with which a student is engaging, and allows the student to see how their solution to a problem plays out in the context. In this way, consequential feedback provides students with information about their reasoning. This paper considers how the timing of consequential feedback impacts the nature of students’ engagement with content.

In general, the literature on feedback has been inconsistent. Feedback can both support (Baron, 1998; Mory, 2004; Sweller, Van Merriënboer, & Paas, 1998) and thwart (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996; Schmidt, 1983) learning. Even when it successfully supports learning, feedback is used in different ways and for different purposes. Much of the research on feedback focuses on general characteristics of feedback, such as timing, specificity, length, and complexity. Feedback is treated as a factor in students’ work, and not as an integral part of their activity. Specifically, the relationship among the content on which students are working, the nature of the activity, and the role of feedback is typically not discussed. Indeed, in her recent review of formative feedback, Shute (2008) notes, “Despite the plethora of research on the topic, the specific mechanisms relating feedback to learning are still mostly murky, with very few (if any) general conclusions” (p. 156, emphasis added). Additionally, research on feedback typically focuses on outcomes such as performance and learning (measured by pre-post change). Although to be sure, there is reason to care about the content that students learn, this is a distal measure that doesn’t yield insight into how feedback is functioning. Understanding how feedback functions therefore requires a closer examination of the nature of feedback as it exists as part of larger activity.

The work presented in this paper builds directly on findings that suggest that what students learn cannot be separated from how they learn it (Barab & Plucker, 2002; Boaler, 2000; Brown, Collins, & Duguid, 1989; Cobb, Stephan, McClain, & Gravemeijer, 2001; Greer, 1991) and that knowing is an interaction among components of complex systems. This shift in the conceptualization of learning highlights the importance of attending to the practices associated with engaging content as part of the learning process, rather than focusing on learning outcomes as a means of determining effectiveness (Greer & Gresalfi, 2008). Building on this framework, the research presented in this paper speaks directly to this gap in our understanding of the design of feedback by considering how a particular form of feedback impacts the nature of the activity with which students engage. This work also speaks to the need for different methodological approaches for studying feedback by leveraging a design-based research methodology that looks iteratively at the nature of activity as it is impacted by design. In what follows, we define and justify the intervention studied in this paper, consequential feedback. We then share findings from two rounds of implementation of design-based research, and discuss the implications of those findings for future research.
Consequential Feedback

Consequential feedback is a form of feedback that is embedded in the context (mathematical or narrative) with which a student is engaging; the way the students’ solution to the problem plays out gives the student information about their reasoning (Gresalfi, 2011). Consequential feedback is elaborated and task level in that it occurs in the context of the problem that is being completed and focuses on process (as opposed to outcome). However, rather than informing students about the mathematical accuracy of their work, consequential feedback offers agency for students to consider whether their mathematical reasoning resulted in the outcome they had envisioned (positioning students as being responsible determining what might have gone wrong). In this way, consequential feedback is more facilitative than directive (Black & William, 1998). Any mathematical task could afford consequential engagement as long as the feedback takes place in terms of the context (rather than in terms of the calculations). For example, if the task is to figure out how long it would take for two trains to pass each other given particular starting points and rates of speed, consequential feedback would allow students to “see” the actual scenario based on their mathematical calculations, through some sort of simulation or engagement in a virtual world (Nathan, Kintsch, & Young, 1992).

Creating opportunities for students to engage consequentially often involves embedding a mathematics problem in a “real world” situation. The context of these problems serves a purpose beyond simply providing “relevance” for disciplinary work. Instead, when well designed, contexts can push back on students’ understanding by forcing the learner to consider the usefulness, impact, or significance of particular tools on outcomes (Cobb, McClain, & Gravemeier, 2003). For engagement to be truly consequential, it must therefore involve engaging procedurally with rules and formalisms, and conceptually by considering why particular solution paths are sensible (Gresalfi & Barab, 2011; Gresalfi, Barab, Siyahhan, & Christensen, 2009). Thus, consequential feedback has the potential to support opportunistic use and meaningful application (not procedural replication) of content (Carr & Claxton, 2002). This way of engaging mathematics supports meaningful disciplinary understanding, as actions cannot have power unless they can be legitimately shown (through proof or defense) to impact particular situations.

Methods

The data reported for this paper come from two iterations of a design-based research project, which sought to better understand how consequential feedback could be integrated into a project-based mathematics unit situated in a virtual educational game called Quest Atlantis. As a design-based research project, our goal was to design an intervention (consequential feedback), which, we conjectured, would support particular forms of mathematical engagement, which we call consequential and critical. The two rounds of data collection presented here, however, are best conceptualized as a quasi-experiment, in that ultimately we compared two iterations of the same curriculum across very similar groups of students. However, these iterations were not randomly assigned; version 1 of the curriculum was used in year 1, and version 2 of the curriculum was used in year 2. These versions were not used in both years because the second version of the curriculum was developed in response to findings from year 1.

Intervention

This study examined consequential feedback through a mission within the online immersive game Quest Atlantis (www.questatlantis.org). The narrative of the mission is based on a well-known project-based mathematics activity from the Adventures of Jasper Woodbury, called “Adventure at Boone’s Meadow” (CTGV, 1997). When students enter Quest Atlantis, they are given an online persona (an avatar) who navigates the virtual world, explores the environment, interacts with non-player-characters, and makes choices to resolve complex real world dilemmas. Thus, the world itself is a part of the information they can explore, and that world can change in response to particular decisions students make. The mission used here informs students that an endangered eagle has been shot in Boone’s Meadow, which cannot be reached by car and takes 10 hours to hike by foot. Luckily, they meet a non-player character who has an ultralight flying machine. Students must decide which route to take, the length and time of the journey, how much gasoline will be required (and where to stop to get it), who will pilot the plane, and whether any additional cargo is necessary (or possible) given the weight limit of the small aircraft. Consequential feedback is built into this narrative by showing students the consequence or outcome of their choices—in particular, whether the flight was completed successfully (without crashing) and whether the eagle survived. The outcome is shown through an image (of the crashed ultralight) or through dialogue with a non-player character. For example, students who make a mathematical error (such as not calculating the amount of fuel needed accurately) see their ultralight crashing, and then learn from the veterinarian that the eagle has died because no one was able to reach her in time. Students who offer a possible but non-optimal solution (i.e. a solution that will take too long) successfully complete their rescue mission to discover that the bird’s wing has been amputated. And finally, students who select both an accurate and efficient route are able to save the eagle. Examples of the kinds of feedback that students experience can be seen in Figure 1.
Study Design
Among other questions, the project considered how the timing of consequential feedback impacted students’ engagement with content. For this paper, we present data two rounds of implementation across two years. Sixty-five students participated in this study. Year one included a 6th grade class of seventeen students, taught by Ms. Bell, and a 7th grade math class of sixteen students, taught by Ms. Kent. Year two included Ms. Bell’s eighteen 6th grade students and Ms. Kent’s fourteen 7th grade students. Both teachers have over five years of teaching experience and have both used Quest Atlantis for several years. Both teachers taught in schools in a suburban part of the Midwestern United States. In year 1, students were given consequential feedback on their strategy after they had invested time in their mathematical calculations (i.e. after they turned in their solution, but before the final reflection). In year 2, students were given consequential feedback based on their initial guesses, before they had engaged deeply with the mathematics. In every other way the curricular experiences were identical: they had the same immersive experience, were given the same information, saw the same possible consequential feedback, and made a recommendation in response to the same prompts. The only curricular difference was whether students received consequential feedback based on their initial solution before making a formal recommendation, or after. It is important to note that these implementations were not “controlled,” in that the same teacher was using the same curriculum in year 1 and year 2. Thus there could be some kind of effect of experience on the findings from year 2. We suspect this is not the case as both teachers needed a refresher about what the activities were about and the mathematics that was covered in the activities. Additionally, the students differed between year 1 and year 2, and although there were no systematic differences, it is certainly the case that individual students can dramatically change the dynamic of a classroom.

Data Sources
Data for this paper draw from several resources: videotapes of whole-class discussion, videotapes of small group discussion, and submitted work that outlined the final recommendation. Videotapes were collected from each day of the implementation (approximately five hours of classroom time from each classroom in both years, a total of approximately 20 hours of videotape). During whole-class time, cameras were set up to capture the entire class’ discussion while focusing on the written work presented in the public space. During small group time, several cameras were placed around the room in order to capture a range of discussions. Groups were chosen randomly based primarily on where students were located in the classroom. Students’ submitted work was in the format of a written recommendation in response to targeted questions about how to best serve the bird (this was submitted before getting feedback in Year one, and after getting feedback in Year two. The recommendation was identical across both years, and included the following information:

“Your task is to give Larry a plan for how to get the eagle back as quickly and safely as possible (in case something like this happens again). Larry will need to know:

• What's the quickest way to transport an eagle from Boone's Meadow to Cumberland City? (include the route, what you need to bring, and who should fly)
• How long will it take?

_TIP:_ Decide what the outcome and travel time would be for multiple plans, and then decide which is best. There isn’t one right answer, but some plans may be better than others. It's up to you to decide which plan is best.”

**Analysis**

The first round of analysis focused on the submitted recommendations as primary indicators of students’ reasoning about the scenario and their understanding of the mathematical content. The recommendations were coded using both a priori and emergent codes. A priori codes included indicators of the accuracy of students’ submitted work (instances where mathematics that was “unreasonable” were noted), and their use of mathematical or consequential justification. Additionally, because a key interest in the study was whether students were engaging critically with content, codes that had been developed to capture these forms of engagement were used (Gresalfi & Barab, 2011; Gresalfi, Barab, Siyahhan, & Christensen, 2009). Additionally, researchers perceived emergent patterns within the students’ submissions, and sought to quantify those differences through the other codes included below in Table 1. Two researchers discussed the codes and further developed more explicit definitions and coding rules. The two researchers separately coded a sample of the Quests (approximately 23% of the total) until 88% of their codes were in common. After confirming 88% reliability, the two researchers independently coded the remaining Quests. Codes were applied in terms of instance. For example, each unreasonable mathematical claim made, or each distinct justification using consequences, was coded as an instance.

The second round of analysis included looking in depth at both whole-class and small-group discussion. These videotapes were examined primarily to document the types of activities that occurred in the classrooms, measuring the duration of whole-class and small group discussions. In addition, the nature of the conversations during whole-class discussions were coded in terms of the amount of time where either mathematical content or narrative content was discussed.

**Table 1. Coding scheme**

<table>
<thead>
<tr>
<th>Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematically unreasonable</td>
<td>The numbers mentioned in the quest show a significant misunderstanding of the mathematics involved. <em>Something that is mathematically inaccurate but reasonable is not included in this category.</em></td>
</tr>
<tr>
<td>Narrative immersion</td>
<td>Expression that student feels like a character in the game, or displays immersion in the narrative of the game.</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>This is aimed to capture when students are imagining what might or should happen, rather than being sure of the consequences of their plan. There is a sense of guessing, a lack of certainty.</td>
</tr>
<tr>
<td>Certainty</td>
<td>Noticeable expression of certainty in the consequences of their plan, such as ‘I KNOW this will work’ (rather than I THINK). This shouldn’t be used to show lack of tentative language.</td>
</tr>
<tr>
<td>Critical engagement</td>
<td>Shows acknowledgement of other options. The student expresses that they had a choice, and chose one thing over another.</td>
</tr>
<tr>
<td>Mathematical justification</td>
<td>Using mathematics to justify their decision, for instance, justifying route (miles, distance), justifying cargo (weight), justifying pilot (weight), justifying time (time), describing one static mathematical claim rather than a causal statement or connection.</td>
</tr>
<tr>
<td>Consequential justification</td>
<td>Using consequences of a plan to justify it (in addition to OR opposed to mathematics), such as Justifying route, justifying cargo, justifying pilot, justifying time WITH an event, NOT A NUMBER, using causal statements.</td>
</tr>
</tbody>
</table>

**Results**

We found that offering consequential feedback in the course of problem solving as in year two led to significantly higher levels of consequential justification (i.e. considering the relationship between mathematics and the story), with students including justification that related to the consequences of their decisions at a rate of five times higher than in year one. Likewise, students in year two included many more instances of mathematical justification (Gresalfi, 2011; see Figure 2) in year two, with the rate of mathematical justification more than doubling. Interestingly, the incidence of mathematical unreasonableness was quite low across both years, indicating that although students might have had some calculational errors, they were primarily engaging with the mathematics in a way that indicated their understanding of the content. Thus, it is not the case that mathematical justification increased because students’ understanding of the mathematical procedures increased.
Likewise, students’ narrative immersion was quite high in year one and in year two, suggesting that the context was equally vivid for students regardless of whether they had received feedback before or after they made their recommendation. This suggests that it is not students’ immersion in the narrative (how “real” or “important” it seemed to be) that supported increased consequential engagement. Finally, students’ critical engagement increased significantly, indicating that students realized that they had some kind of choice about how they might solve the problem. This difference in critical engagement might give insight into the differences in justification, as discussed below.

Figure 2. Coded Final Recommendations Collapsed Across Classes For Year 1 and Year 2

The trends captured by collapsing the classes reflect what happened in the individual classes, although looking between the classes an interesting difference can be seen. In comparing Ms. Bell’s class from Year 1 and Year 2, differences are apparent in the codes of critical engagement, consequential justification, mathematical justification (see Figure 3). Submitted recommendations from Year 2 displayed over six times more critical engagement than year one, nearly five times more consequential justification, and about three times more mathematical justification. In comparing Ms. Kent’s class from Year 1 and Year 2, similar differences are apparent in the codes of critical engagement and consequential justification; students’ submitted recommendations in Year 2 showed about four times more critical engagement, and six times more consequential engagement than students’ submitted recommendations in Year 1. Interestingly, the trend did not apply for the code mathematical justification, which showed no change.

Figure 3. Coded Final Recommendations For Ms. Bell and Ms. Kent for Years One and Two.
The coding of the final recommendations suggest that embedding feedback in advance of submitting final recommendations supports students to include more justification in their responses. This is important as students’ use of justification gives teachers greater insight into their students’ thinking (important as an aspect of formative assessment). In addition, justification creates opportunities for students to think more deeply about their own mathematical ideas, and thus more and better justification likely supports deeper understanding of mathematical content (Niemi 1996). However, these quantitative results give no indication of why the change in timing of consequential feedback led to differences in justification, nor why changes in mathematical justification occurred in one classroom and not the other. In seeking to better understand students’ activity, we examined classroom discussions in more detail, characterizing both the nature of the activity and the content of the conversations that occurred in the two classrooms over the two years.

Further Investigation

As discussed above, the ratios of change for each category were nearly the same, with the exception of mathematical justification. In one class (Kent), the codes for mathematical justification did not change at all, while in the second class (Bell) increased by over four times its original number. The change is particularly interesting when considering that all classes except for Bell Year 2 had mathematical justification totals of 14 or 15. The comparisons between the classrooms across two years allow us to eliminate some potential hypotheses. First, differences in the teaching styles of Ms. Kent and Ms. Bell would not (alone) have been likely produce the increase in mathematical justification for Ms. Bell in the second year. Any teaching differences would have also have been likely to impact the students’ submitted recommendations for year one as well.

One contributing factor might be the time the frequency of whole class discussions—perhaps Ms. Bell simply talked more with her students, thus giving them more opportunities to understand justification and its purposes. As shown in Table 2, whole class discussion time varied among the four environments observed. Discussion in Ms. Kent’s class remained nearly the same while discussion in Ms. Bell’s class increased dramatically in Year 2, considering both amount of time and frequency of discussion. This presence of more frequent discussions, and a very large overall time in discussion, occurred only in Ms. Bell’s Year 2 class, and therefore, could account for some difference present within that class.

Table 2. Amount and percent of time spent in whole class discussion

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of Discussions</th>
<th>Average Time In Each Discussion</th>
<th>Time in Discussion Total</th>
<th>Percent of total class time in discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bell Y1</td>
<td>2</td>
<td>19.45 minutes</td>
<td>47.5 minutes</td>
<td>26%</td>
</tr>
<tr>
<td>Bell Y2</td>
<td>6</td>
<td>10.6 minutes</td>
<td>69.5 minutes</td>
<td>52%</td>
</tr>
<tr>
<td>Kent Y1</td>
<td>2</td>
<td>15.9 minutes</td>
<td>31.7 minutes</td>
<td>13%</td>
</tr>
<tr>
<td>Kent Y2</td>
<td>3</td>
<td>9 minutes</td>
<td>27 minutes</td>
<td>8%</td>
</tr>
</tbody>
</table>

One interesting note is that although the total amount of discussion time increased in Ms. Bell’s class in year two, the actual time discussing mathematics was almost identical. Indeed, the actual percent of the discussion that focused on mathematics decreased significantly for year two. Thus it is likely not the case that mathematical justification increased for students in Ms. Bell’s class in year two simply because they had received more instruction about mathematics.

Table 3. Amount of time spent in mathematical and consequential discussion

<table>
<thead>
<tr>
<th>Class</th>
<th>Time in mathematical discussion</th>
<th>Percent of total discussion time</th>
<th>Time in consequential discussion</th>
<th>Percent of total discussion time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bell Year 1</td>
<td>24 minutes 32 seconds</td>
<td>52%</td>
<td>103 seconds</td>
<td>4%</td>
</tr>
<tr>
<td>Bell Year 2</td>
<td>24 minutes 50 seconds</td>
<td>36%</td>
<td>225 seconds</td>
<td>5%</td>
</tr>
</tbody>
</table>

Considering the evidence that the discussions in Ms. Bell’s class did not change drastically in their content, it is still plausible that having more time in discussion, and having more frequent discussions may have contributed to the sensitivity of students to the consequentiality factor. The mechanism of the contribution may not be as simple as having more mathematical or consequential discussion, but perhaps the timing of the discussion, or the ways in which discussion was structured.

In addition to instructional factors that may have impacted the effects of consequentiality within the game, an additional conjecture might be that there is a difference between the two classes with respect to students’ sensitivity to consequential feedback. The two classes were very different environments and had quite different student populations. In addition to individual differences of students, the Kent classes were in 7th grade...
within a project based school, while the Bell classes included 5th grade students enrolled in 6th grade math in an elementary school. It seems possible that the younger students were more likely to “play” in the narrative of the storyline, thus actually engaging the complexity of the situation more deeply. Indeed, in examining small group interactions, there were several conversations amongst students in Ms. Bell’s class wherein they seemed to be deeply immersed in the narrative, seen in the ways they played out scenarios beyond what had actually happened (using their imagination to fill in the gaps that the videogame had left). For example, in a small group interaction wherein students were reviewing the feedback they had received, one student summarized his experience as follows:

“The first time I didn’t really think about it. I just said, this time will be… I’m just going to play around with it. So, the first time I tried it, I took Route Two. It was straight up to Boone’s Meadow and straight back. Um, I took the extra tank of gas, and I didn’t take the cargo box because I thought it would overload the payload. And, I thought, um, basically (laughs)… because when I got there, the eagle had to obviously ride somewhere. And, where was it going to fit? It was a one-person seat, so, um (pause) the eagle fell out (laughs), and I ran out of gas. I found the eagle and took it back to the vet, um, but it died.”

Several pieces of this student’s explanation were not included in the actual experience: the eagle can’t fall out of the ultralight, and if the student ran out of gas, they experienced the ultralight crashing and then immediately arrived at the door of the vet (they did not have an opportunity to find an eagle after the crash). Thus, in recounting his experience to his two classmates, this student was introducing new details that made his narrative consistent, thus immersing himself deeply into the context. Likewise, students in this class sometimes spontaneously offered factors that might impact the bird’s survival beyond the information given in the scenario. For example, the following solution was offered as an alternative suggestion (note the consequential justification):

K: I was going to say. The reason I chose going to here is that it might be a little easier on the eagle because a) there would be no predators
T: Because they couldn’t get in the plane?
K: No outside bacteria or germs or it couldn’t get shot again, which is kind of like still (a possibility). And one more thing. You could immediately give first aid to the eagle if you pick it up.

In both of these instances, students were immersing themselves in the “real world” of saving the eagle based on their experience of playing out the mathematical solutions they had proposed. This kind of imaginative conversation, wherein students were generating their own information and examples, was only seen in a single utterance in Ms. Kent’s classroom.

Discussion
This paper has presented the results of two implementations of a curricular activity that considered how the timing of consequential feedback might impact students’ engagement with mathematics. Findings suggest that when students have an opportunity to experience consequential feedback before making formal recommendations, their solutions include much higher rates of justification, both mathematical and consequential. These findings are particularly promising given that they bore out across implementations in two very different classrooms. It is not clear why including feedback in advance of careful mathematical calculation, as opposed to after completing these calculations (when feedback typically occurs) is so significant in terms of impacting the ways students engage mathematically. One hypothesis is that locating consequential feedback in the midst of students’ problem solving is truly formative in that the feedback helps students gain insight into the impact of their mathematical calculations. Consequential feedback can become another data point against which students can contrast their mathematical decisions. For example, in a recent implementation, two students discussed two different outcomes—one solution that led to an eagle having his wing amputated, and another that allowed the eagle to live unscathed. These contrasting outcomes prompted the students to re-calculate the time it would take to complete a targeted route, which involved engaging in the mathematics of ratio. Thus, one hypothesis of this study is that consequential feedback actually prompts students to engage more deeply with mathematical content because it prompts them not just to correct, but also to think about how the mathematical works in terms of leading to solutions that aren’t desired. Indeed, the awareness of possible different outcomes might also explain why instances of justification increased; students realized that there was something that actually had to be justified.

However, specific differences between the two classrooms, in particular with respect to mathematical justification, led to deeper investigation into the nature of the implementations across these two classrooms. As detailed above, we had several conjectures about what might have contributed to these differences, some of
which were not borne out. Specifically, it wasn’t the case that the classroom that had higher rates of mathematical justification simply talked more about math; nor what it the case that the teacher in that classroom gave different instructions that emphasized mathematical justification. Differences in the total amount of classroom conversation might have contributed to these differences. Alternatively, something about the nature of students’ engagement with the narrative—and the extent to which that helped them become aware of the myriad possible solutions—might also have contributed to this difference. The differences between the nature of students’ engagement with the narrative suggests a possible avenue for future research: considering to what extent narrative immersion becomes a resource for students’ meaningful conceptual engagement.

References