

CASE STUDIES

Turning the King Tide: Understanding Dialogue and Principal Drivers in an Online Co-Created Investigation

Karen Peterman*, Christine Bevc† and Ruth Kermish-Allen‡

Online learning communities for citizen science have been growing in number and scale in recent years. The *WeatherBlur* project was designed to apply knowledge-building theory to a non-hierarchical online citizen science community for students in grades three to eight, their teachers, and scientists. This case study explores one investigation to determine the kinds of interactions that encouraged online knowledge-building and the individuals who served as the key drivers for the investigation. Posts from the project's discussion board were analyzed via a discourse analysis. The results indicated that students, teachers, and scientists initiated online discourse at similar rates, affirming the project's non-hierarchical design. Two knowledge-building constructs were used to initiate dialogue, including comments that added to the quantity and then quality of information in the conversation and those that included some conceptual advancement. Photos were also used more often to initiate rather than respond to conversation. By comparison, ongoing conversation was sustained by comments that enhanced the quality of information shared. Using activity log data, the results from a social network analysis indicated high variation in posting frequency, high probability of reciprocity, and dense core-periphery structure. Both the discourse analysis and social network analysis revealed multiple ways for individuals to become core to the network and thus primary drivers for the investigation. The findings are discussed in relation to knowledge-building in online citizen science communities.

Keywords: online; community; discourse; knowledge-building; network analysis

Introduction

Technology increasingly plays an important role in citizen science, with online tools being used to house learning protocols, submit data, and share results and interpretations (Connors, Lei, and Kelly 2012; Crall et al. 2010; Newman et al. 2010). Many citizen science projects strive to create learning communities in which individuals come together online to work toward the common goal of building scientific knowledge. Discussion boards are a common element of study for these types of projects. Research has begun to document the networks and interaction patterns of volunteers who participate in these online spaces (Aristeidou, Shanlon, and Sharples 2015; Curtis 2015; Luczak-Roesch et al. 2014; Tinati et al. 2014). Most citizen science discussion boards were created to allow community members to ask and answer each other's questions (Luczak-Roesch et al. 2014). Projects that encourage volunteers to conduct the full inquiry process via the discussion board are less common, though some examples do exist (Aristeidou et al. 2015, 2017).

This case study illustrates an approach to creating online citizen science communities by adopting the knowledge-building community framework. "Knowledge-building is a social process through which people work collaboratively to create and improve ideas of value to their community" (Sun et al. 2010; p. 148). Knowledge-building communities are characterized by an overall goal of advancing knowledge beyond what is currently known (Scardamalia and Bereiter 1994). A key difference between online communities that engage volunteers via a typical discussion board and a knowledge-building community is the fact that the latter carries an expectation that problem-solving and meaning-making will occur primarily through online collaboration and discourse. Successful knowledge-building communities require a number of characteristics: 1) attempts to understand the world, starting with personal experience and observation; 2) a shared commitment among community members to generate coherent and usable new knowledge; 3) a shared responsibility across the community for collaboration; 4) a willingness to negotiate ideas to advance understanding; 5) the ability to build upon and be critical of past knowledge, ideas, and artifacts; and 6) knowledge-building discourse, in which participants engage in constructing, refining, and transforming knowledge (Hmelo-Silver and Barrows 2008; Zhang, Scardamalia, Reeve, and Messina 2009).

* Karen Peterman Consulting, Co., US

† RTI International, US

‡ Maine Mathematics and Science Alliance, US

Corresponding author: Karen Peterman
(karenpetermanphd@gmail.com)

When transferred to a classroom, knowledge-building discourse is grounded in issues that are relevant to students' personal lives (Scardamalia and Bereiter 1994; Bereiter and Scardamalia 2003). Three types of support are needed for classroom-based knowledge-building discourse: (a) a focus on problems and depth of understanding; (b) decentralized, open knowledge environments for collective understanding; and (c) productive interaction. The *WeatherBlur* project applied these concepts via an online citizen science community that allowed students, teachers, scientists, and community members to identify issues that were related to local weather or climate problems. Investigations were designed as co-created citizen science projects that empower participants to utilize the entire inquiry cycle, from generating research questions to communicating results. Using the program's online discussion board, participants contribute to each topic by building from their own local, historical, and scientific expertise. They then work together to gather data and generate new knowledge on each investigation topic.

This case study bridges the field of citizen science and that of knowledge-building communities by documenting the discourse and principal drivers of a citizen science investigation using both discourse and social network analyses. Discourse analysis involves the systematic coding of communication. Within much of the knowledge-building literature and within the context of this study, the discourse consists of the written contributions posted to the online community. Social network analysis is a technique that uses relational data to identify patterns in social structures, such as those generated through the discussion boards of online learning communities. The next sections detail prior research that has used each of these analytic techniques to document knowledge-building and online citizen science participation, respectively.

Knowledge-building as described by discourse and social network analysis

The relationships among the knowledge found in discourse are "the most important indicators of learning activity" for a knowledge-building community (Oshima, Scardamalia, and Bereiter 1996; p. 132). Discourse analysis of knowledge-building projects with upper elementary and middle school students has focused on the types of knowledge connections that students make (Oshima et al. 1996), the conceptual advancement provided by students during online discourse (Hakkarainen and Sintonen 2002), the levels of explanations that students use to build knowledge (Hakkarainen 2003), and the importance of questions (Lai and Law 2013). The evolution of these schemes has resulted in a recent paper that documents "good moves" in knowledge-building, highlighting seven dialogic actions that promote knowledge creation and critical thinking (Bereiter and Scardamalia 2016).

Discourse analysis of online knowledge-building among youth has indicated that even primary school students can work together to build knowledge (Chen, Scardamalia, and Bereiter 2015), and that this process results in enhanced understanding of science content (Resendes et al. 2015). Lin and Chan (2018) found that fifth grade students' problem-centered posts were predictive of

enhanced understanding of epistemic cognition (i.e., the understanding of how people think about the nature of knowledge and knowing in science), while theory-building posts were predictive of gains in understanding science content. Lia and Law (2013) identified a developmental difference between how middle schoolers and high schoolers used questions and explanations to advance knowledge. Though both groups asked sophisticated questions in their online discourse, these questions resulted in higher-level explanations only among older students.

Though the utility of social network analysis to study knowledge-building has been tested and verified (Oshima, Oshima, and Matsuzawa 2012), studies that examine the relationships between students to assess knowledge-building are rare. Using this relational approach, Zhang et al. (2009) found that opportunistic groups that evolved based on a common goal provided the most distributed network structure, such that students engaged with a wider range of participants across the community and at a level similar to that of their teacher. Palonen and Hakkarainen (2013) used social network analysis to understand the interactions between students and found that there were large individual differences in the number of contributions made by elementary school students, and that females facilitated the majority of knowledge building.

Participation in online citizen science communities, as described by discourse and social network analysis

Like the knowledge-building literature, discourse analysis of online citizen science projects is more prevalent than the use of social network analysis to study volunteer interactions. Discourse analysis of discussion board posts has been conducted to learn more about users who participate in citizen science, the relationship between the amount of discourse and data contributions, and whether online discussions help users to feel a sense of community (Aristeidou et al. 2015, 2017; Huang, Cheng, and Huan et al. 2018; Liberatore et al. 2018; Luczak-Roesch et al. 2014; Tinati et al. 2014; Tinati, Luczak-Toesch, and Simperl et al. 2015; Tinati, Luczak-Toesch, and Simperl et al. 2016).

One of the most consistent findings across these studies is that a small number of volunteers do the majority of the work. This pattern is true regarding both the amount of data submitted and for participation in online discussion boards (Aristeidou et al. 2017; Curtis 2015, 2018; Luczak-Roesch et al. 2014). The presence of a "super user" group is such a constant that researchers have used the phenomenon to categorize new projects and to make recommendations about leveraging these groups to help sustain online citizen science communities (Ono, Ikkati, and Enoto 2018).

Of the studies in the citizen science literature, the approach used by Huang et al. (2018) is the most similar to that used by knowledge-building researchers. Examining the scientific discourse of citizen scientists as they engaged in modeling and problem solving online, Huang and colleagues found that the timing of facilitator input was most important in supporting collaboration when they initiated and framed discussions and as they helped to wrap up discussions. In addition, the use

of leading questions and sharing strategies to further the investigation each played a critical role for participants in advancing knowledge related to planning and problem solving. This evidence is similarly found in the network-based analysis of the *Weather-It* project, which demonstrated that some volunteers actively engaged in the project by creating new investigations and data, while others simply commented on the mission and data posts provided by others (Aristeidou et al. 2015). Earlier work examining the relational data from *WeatherBlur* has suggested that the project's online community model was successful at breaking down traditional hierarchies of learning (Kermish-Allen, Peterman, and Bevc 2018). This study extends these findings by answering two research questions:

1. Which forms of discourse encourage knowledge-building in online citizen science investigations?
2. Which users, if any, are the principal drivers of online citizen science investigations?

Methods

The data for this study include discussion board posts as well as server-side data. This research was reviewed by the New England Institutional Review Board and considered to be exempt (NEIRB #12-178).

Study context

WeatherBlur consists of a non-hierarchical knowledge-building community in which participants function as both co-creators of knowledge and active learners via citizen science projects. Working as a community of practice (Wenger 2000), students, teachers, scientists, and community members come together and work toward a common goal, sharing actions and interactions to promote knowledge-building in relation to local citizen science investigations.

There are no predefined investigation topics in *WeatherBlur*. Instead, investigations are formed through online discourse as participants engage in an iWonder process to propose and hone questions that can be defined as a formal investigation. Once the investigation is defined, the community works to design data-collection protocols by building on existing methods, leveraging both the place-based and scientific expertise of community members. Anyone in the online community can then join the investigation to participate in knowledge-building.

The qualitative and relational data for this study were collected as part of an investigation that focused on the November 2016 king tide in the United States, Maine. King tides are defined as the high tides that occur when the moon is closest to the Earth. This investigation topic was initiated by a fourth grade student interested in learning more about how the high water from the king tide would affect local beaches, roads, and buildings in her community. Her post prompted a larger discussion about the potential impact in other communities and how the effects could be measured. A scientist in the *WeatherBlur* community joined the group based on his interest in adding hyperlocal data to the few existing tide monitoring stations up and down the coast which contribute to

a statewide tide-gauge database. Participants worked together during an initial three-week formative discussion period to refine the research question and define the data-collection protocols that supported the formal investigation. The King Tide Investigation was designed to gather three types of data: Photographs and videos of sites before, during and after the king tide; measurements of a high tide before the king tide, along with calculations of mean sea level; and the height of the king tide with some basic wind and other weather information. Student participants often completed these tasks as a class.

Online discourse within the investigation took place over seven weeks, beginning with planning conversations that occurred almost exclusively between teachers and scientists. Then, on the day after the king tide, a staff member asked students to share their observations and findings. Students provided a number of responses, including those posted in response to a class assignment to share data collection procedures and observations. The final contributions to the King Tide Investigation were also posted via a class assignment in which students uploaded data presentations to describe their results.

Participants

Thirty users participated in the King Tide Investigation, including two scientists, one facilitator, four teachers, two community members, and 21 students. Students were in grades 3–5. Five schools participated in the investigation, including three small, island-based schools with grade level sizes ranging from one to 14 students. The two remaining schools were more traditional in size. Most students participated in the investigation as an entire class. The sample for this study includes only students who made online contributions to the seven-week investigation. During that time, participants contributed either online discourse or photos via the online community's discussion board. A total of 53 conversations generated 250 comments.

All online activity was recorded by server-side activity logs, including comments, responses, and photos posted. These served as the dataset for the present study. The activity log data were downloaded by the research team at the end of the investigation and compiled into a series of comma-separated files for coding and analysis. Discussions were identified through a unique numeric "parent" identifier that was used to link each individual comment to prior and related discussion posts; this function allowed the research team to see the growth and flow of discussion over time. Row-separated time-stamped entries of each comment were used to define the unit of analysis for both the discourse analysis and social network analysis. Comments also were clustered into discussion "threads" as a second unit of analysis.

Discourse analysis

A discourse analysis was conducted to replicate the approach used to study knowledge building. Online posts were coded and analyzed using NVivo 11. Two series of codes were of interest. The first series documented the context during which online discourse occurred. The second was used to document how knowledge was shared and

built through the online community. Codes were assigned on a turn-by-turn basis to discussion board posts. Consensus coding was conducted for all posts. Two researchers coded each post independently and then compared codes; all disagreements were discussed and the final code was agreed upon.

Context

Context codes were used to ground the discourse analysis. First, all posts were coded to indicate whether they included *statements* or *questions*. Next, posts were coded as initiations or responses. *Initiations* were defined as the first post related to a new topic that resulted in at least one response. *Responses* were defined as subsequent posts on an existing topic. The presence of *photos* was also coded.

A second series of context codes captured posts that were aligned with the investigation process. *Background knowledge* identified information that was shared to support the investigation; examples included foundational knowledge, information on study sites, and historical data. *Procedure* was used to capture information that shared or clarified plans for data collection, such as calibration of instruments. *Data* indicated posts that featured an observation, data point, or photo captured for the investigation. *Analysis* was used to code posts that summarized investigation data, as well as posts from others that helped to interpret data. A final

code was used in instances for which none of the prior codes applied.

Knowledge-building

Coding related to four constructs was informed by the knowledge-building literature. The full list of codes and definitions is provided in **Table 1**. For longer posts in which multiple codes were applicable for one construct, the highest level code was assigned.

Knowledge connections were coded, using categories developed by Oshima et al. (1996). Four codes were used to document comments that widened the conversation, deepened the conversation, or both (see **Table 1** for definitions). A final code was used for instances in which no knowledge connections were made.

The next construct, *levels of explanation*, was adapted from Hakkarainen (2003). *Levels of explanation* codes were based on the extent to which the information in a post was articulated and organized clearly. *Facts* were considered the least sophisticated of the codes applied, followed by *partial* and *full explanations*. A final code was used to document posts that included no explanation.

Conceptual advancement was captured using three codes adapted from Hakkarainen and Sintonen (2002). The level of conceptual advancement was coded in relation to information about either scientific content or the investigation process itself, and with regard to the post's

Table 1: Knowledge Building Categories and Code Definitions.

Categories and Codes	Definitions
Knowledge Connections	
Knowledge widening	Posts that added a piece of new knowledge to a topic already represented in the conversation, resulting in an increase in the quantity of information in the discussion.
Knowledge deepening	Posts that increased the level of knowledge represented, representing a shift in the quality of information in the discussion.
Widening, then deepening	Posts that added a piece of new knowledge to an existing topic in the conversation, and then increased the level of knowledge represented.
Deepening, then widening	Posts that increased the level of knowledge presented, and then added a piece of new knowledge to an existing topic in the conversation.
No knowledge connections	Posts that did not include knowledge connections.
Levels of Explanation	
Facts	The presentation of information or lists of information, without an explanation.
Partial explanation	Posts that attempted to provide an explanation, but that included limited or only partially articulated information.
Full explanation	Posts that provided a well-elaborated and comprehensive description.
No explanation	Posts that did not include either facts or an explanation.
Conceptual Advancement	
Some Advancement	Posts that included some pieces of new information related to the investigation content or methods, such that the post was likely to facilitate a moderate shift in conceptual understanding for the group.
Strong Advancement	Posts that included substantial pieces of new information related to the investigation content or methods, such as new theoretical concepts or explanatory theories, which were likely to result in a significant advance in the online discussion.
No advancement	Posts that did not include new information with the potential to advance the investigation content or methods.

potential to move the investigation forward. No advancement was also coded, as applicable.

Social network analysis

Social network analysis was used to identify the patterns and interactions of discourse and exchange between users in the online community over time and to identify the principal drivers of the investigation (i.e., those individuals who were most important to the functioning and growth of the network). Each comment in the activity log contained information that uniquely identified who posted the comment (i.e., the sender) and the user of the “parent” comment to which the post referenced (the receiver). This combination of sender-receiver (e.g., Sue responded to Kate’s comment) represents one unit of analysis. The social network was generated from the listing of this information captured in the log. Analyses were conducted in the *statnet* suite of packages in the R software platform (Wasserman and Faust 1994; Handcock et al. 2008; R Core Team 2016).

Once the network was created, the analysis examined not only how individual users participated in the investigation, but also how their participation contributed to the larger discussion and the discussions of other users. The data were first analyzed to identify variations across individual users, evaluating the number and types of connections they had with others. Two primary measures were used to document the number of comments that users made and received (*degree centrality*), and how often users were placed in a position between two other users and thus with the potential to serve as a “go-between” (*betweenness centrality*) in the discussions (Freeman 1978).

The second part of the analysis measured the interactions and patterns of exchange between two or more users, reporting on the *reciprocity* and *core-periphery structure* of the network (Wasserman and Faust 1994). *Reciprocity* measures the mutual exchange in the knowledge-building process. For example, if Sue posts a question and receives a comment from Kate, reciprocity reports how often Sue responds to Kate to complete the social interaction. This reflects an incremental measure of knowledge-building, whereby reciprocity is only one part of the larger process.

The *core-periphery analysis* was used to differentiate the various “layers” of the network to identify principal drivers. To uncover the “layers,” the analysis calculated k-index (k_i) scores to provide an individual measure of a user’s position relative to the “core” or center of the investigation. The k-index reflects a user’s aggregate number of connections, centrality in the network, and “role” as go-between in relation to others, relative to size of the network. This measure was calculated for each user and the user’s scores were then ranked. The greater the index value, the more critical the role a user played in the overall network. Relative to network position, the results of this analysis offer a more nuanced understanding of the individual and collective dynamics contributing to user discussions.

Results

The discourse and network data were analyzed in parallel. This section presents each analysis sequentially and then in an integrated format to document the collaboration

and knowledge-building processes that occurred within the investigation.

Research question 1: Which forms of discourse encourage knowledge-building in online citizen science investigations?

To answer the first research question, trends were compared between discourse that was used to initiate new conversation versus discourse that responded to existing conversation. The totals for each set of codes, along with the results for chi square tests that compared the portion of instances used to initiate versus respond to conversation, are presented in **Table 2**. A Bonferroni procedure was used to control for Type 1 error; results were considered significant at the .003 level.

A total of 53 conversations were initiated during the investigation. An additional 197 posts were made during the investigation. Most were coded as *responses* ($n = 150$; 75%) resulting in a total of 203 comments that were compared in the analysis. As shown in **Table 2**, *background knowledge* was the context coded most often, followed by *data*. The only context used differently across initiations and responses was *analysis*, which was used more often to initiate than to respond to conversation.

Knowledge connections seemed to function in distinct ways depending on whether a comment was an initiation or a response. *Knowledge connections* to *initiate* a conversation were significantly more likely to *widen then deepen* the conversation, while *responses* were significantly more likely to *deepen* the conversation only. For example, the photo in **Figure 1** was presented by a student who initiated conversation with the following *widening, then deepening* post: “We graphed the high and low tides for a month before the king tide. We noticed the tides made a pattern. It kind of looks like a mountain.” In response, a scientist *deepened* the conversation by saying, “This is an excellent graph.... The camera angle is good too because it allows the eye to look at the trends easily. You might make an overlay of the phases of the moon to see if that helps explain the bumps and valleys. Your explanation can also talk about why the width of the blue band varies.”

The *level of explanation* provided in posts did not vary significantly across *initiations* and *responses*. *Facts* were coded most often, followed by *full* and then *partial explanations*. Differences were found in the use of *some conceptual advancement* to initiate rather than respond to conversation. The following post from a teacher provides an example of an initiation with *some advancement*: “Check out the photographs taken by [a] drone over Brown’s Boatyard, North Haven, Maine, during the King Tide on Tuesday, November 15, 2016. There are more under observations.”

Photos were also used significantly more often to initiate rather than respond to existing conversation. **Figure 2** shows a photo of a map that was posted by a scientist in preparation for the king tide event, with the following comment: “Here comes a low pressure center (L) up the coast from the south. It is not a very big storm but could push a little more water ashore during the king tide.”

Table 2: Use of Context and Knowledge-Building Codes to Initiate and Respond to Online Discourse.

Categories and Codes	Instances Coded	Initiations	Responses	χ^2
Context				
Background knowledge	84	39%	42%	ns
Procedures	20	7%	11%	ns
Data	49	17%	27%	ns
Analysis	27	30%	7%	17.01***
Knowledge Connections				
Widening	20	20%	6%	ns
Widening, then deepening	59	57%	9%	52.95***
Deepening	115	6%	38%	19.72***
Deepening, then widening	57	9%	17%	ns
Levels of Explanation				
Facts	101	35%	28%	ns
Partial explanation	33	24%	7%	ns
Full explanation	61	17%	17%	ns
Conceptual Advancement				
Some Advancement	114	56%	28%	12.97***
Strong Advancement	41	17%	11%	ns
Photo included		53%	12%	38.59***
Questions included	51	19%	28%	ns

*** $p < .001$.

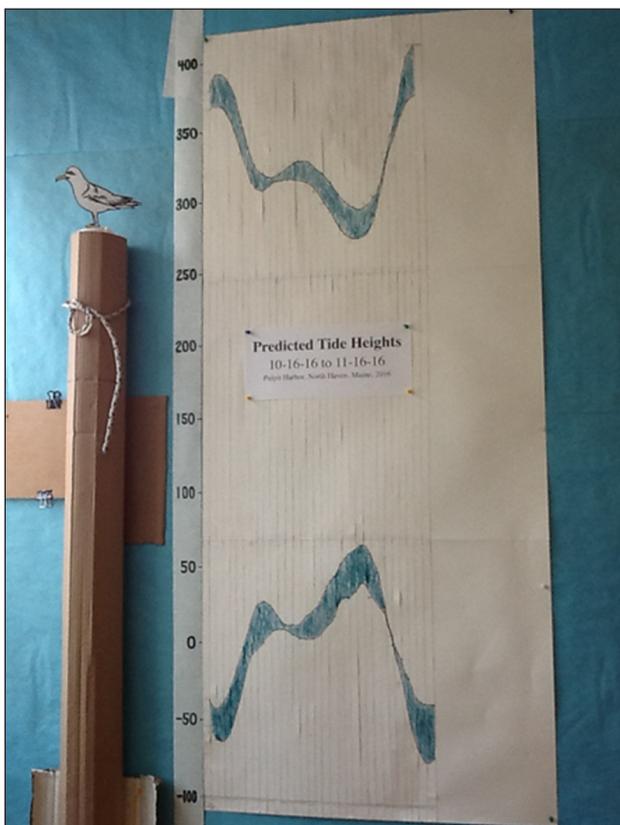


Figure 1: Sample of student graph used to widen and then deepen online dialogue.

The use of *questions* did not vary across initiations and responses. When questions did occur they did not seem to advance the conversation; fewer than half resulted in an answer (44%). Most questions asked for a simple clarification such as the following example: “This looks great. Where is it located?” to ask about a water gauge. Explanation-seeking questions also were common, such as the following question from a teacher, “I wonder how the predicted wind is going to affect the zero tide tomorrow. We will still try to calibrate our gauge at 1:37 PM, but I am wondering if the reading should be considered suspect. Any thoughts?” In summary, given that *initiations* were defined as new topics that yielded at least one response, conversation that focuses on *analysis*, the use of *widening, then deepening* comments, and providing *some advancement* can each be considered strategies that encourage new online conversation. *Deepening* comments offer a strategy for sustaining conversation once initiated.

Research question 2: Which users, if any, are the principal drivers of online citizen science investigations?

The results presented thus far have focused on how different types of discourse functioned within the context of the King Tide Investigation. This section focuses on the participants who contributed those responses. Given the small sample size and uneven distribution of sub-groups, statistical analyses were not conducted by group for the

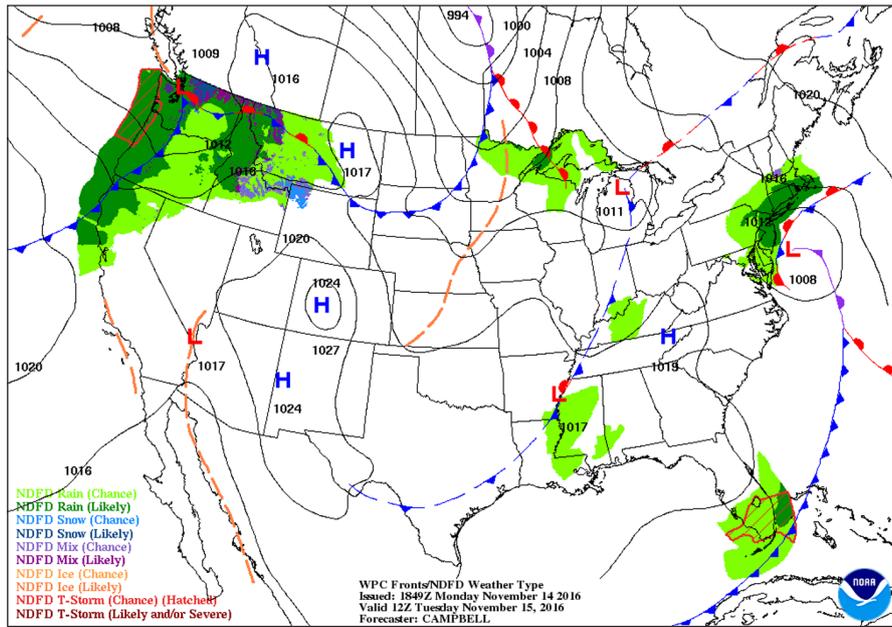


Figure 2: Sample photo of a weather map used to initiate an online conversation.

discourse analysis. Percentages are reported throughout for descriptive purposes.

First, differences in the King Tide Investigation were explored by user group. Online conversation was initiated most often by students (42%), followed by teachers (36%), and then scientists (22%). Similar portions of the responses were posted by scientists and students (38% and 36% of responses coded, respectively); slightly fewer were posted by teachers (26% of all responses).

Looking at the context of online dialogue, each user group dominated different portions and thus played a distinct role in the investigation. Over half of the *background knowledge*, for example, was provided by scientists (55%). Scientists were also most likely to ask *questions* (60% of the instances recorded). Discussion of *procedures* was dominated by teachers (54%). Teachers also submitted approximately half of the *photos* that were posted to the online discussion (52%). Teachers and students were equally likely to share *data* (43% and 41%, respectively). Students provided the majority of the data presentations that were coded to *analysis* (59%).

With regard to knowledge-building constructs, user groups differed slightly in their overall use of *knowledge connections* (**Table 3**; overall constructs are in bold text). Scientists and teachers made a similar portion of the contributions to this construct, and at a higher level than students. The use of specific *knowledge connections* varied by user group. Over half of the *widen* responses were from teachers, and students contributed more than half of the *deepen, then widen* posts. Scientists contributed the greatest portion of *deepen* comments, and *widen, then deepen* comments were provided at equal levels by scientists and teachers, and in greater portion compared to students.

The largest portion of *levels of explanation* was provided by students, with scientists and teachers providing similar lower portions. All groups contributed a similar portion of *facts*. Scientists and students provided a greater portion of both *partial* and *full explanations* compared to teachers.

Table 3: Group Differences in Knowledge-Building Codes.

Categories and Codes	Scientists	Teachers	Students
Knowledge Connections	38%	35%	27%
Widening	33%	52%	14%
Widening, then deepening	36%	36%	28%
Deepening	44%	33%	23%
Deepening, then widening	35%	13%	52%
Levels of Explanation	30%	27%	43%
Facts	30%	35%	32%
Partial explanation	33%	19%	46%
Full explanation	35%	21%	39%
Conceptual Advancement	41%	33%	26%
Some Advancement	34%	43%	24%
Strong Advancement	61%	7%	32%

Scientists provided the greatest portion of comments that promoted *conceptual advancement* overall, followed by teachers, and then students. Teachers contributed the highest portion of comments coded as *some advancement*, followed by scientists, and then students. Scientists provided the majority of comments coded as *strong advancement*, followed by students.

Though these results show some variability by user group, the number and type of contributions seem similar overall. Students and teachers, for example, were similar in the number of *initiations* provided, and students and scientists were similar in the number of *responses*. There was also variability by group in the knowledge-building constructs used, but these differences were not based on the complexity of the discourse. These findings provide evidence to support the non-hierarchical design of the online learning community overall.

Next, social network analysis was used to document the importance of different users and groups to the King Tide Investigation. Each user had an equal opportunity to post to the online discussion space. In **Figure 3**, each symbol represents a different user and the shape reflects a different user type. The size of the symbol reflects the total number of comments posted (i.e., sent) by each user, and the arrowed line represents who commented to whom, including self-responses (loops). Collectively, this network diagram illustrates the exchange of posts and comments between users on the discussion board.

As the figure shows, several users were very active, posting frequently; the top three users included one participant from each user group: Scientist A (#5 in **Figure 3**) posted 55 times, Teacher B (#15) made 45 postings, and Student C (#21) made a total of 21 comments within the discussion (*out-degree centrality*). These values differ from the number of comments or replies received by users (*in-degree centrality*). For example, Scientist A received 34 responses (*in-degree*) versus the 55 comments made as part of the investigation discussion. In contrast, some users received more comments than they posted, e.g., Scientist B (#4) received 18 responses with respect to her 11 posted comments.

The mutual exchange of information reflects another critical aspect of the knowledge-building process. As the arrows in the figure show, many interactions were one-way and not reciprocated. Of the discussions initiated by users in this investigation, 35% were mutual exchanges versus 39% that were one-way (asymmetric) and 26% that were self-responses. The overall measure of *edge-wise reciprocity* among users was $r^1 = 1.568$, which can be interpreted as the relative log-odds of a discussion post or comment given a reciprocation, versus the baseline

probability (50/50) of a post or comment occurring in the network (Butts 2008). In other words, where all users have an equal opportunity to respond to posts, a comment by User A to User B is about 56% more likely to receive a response from User B than other users in the discussion space, “closing the loop” on the discussion and helping to build the investigation.

On average, *betweenness centrality* was 16.17, with a median of 0.23 (interquartile range of 0 to 6.93), indicating that many users were frequently positioned in the network between other users. This position contributes to the user’s potential to serve as a broker, or go-between, for other users based on his or her position in the network. The interactions between users are also important when considering the exchanges that contribute to the building of knowledge. The users found to have the highest *betweenness centrality* scores were Scientist A (#5, $C_b = 183.7$), Teacher B (#15, $C_b = 94.0$), and Student C (#21, $C_b = 55.2$), the same users who posted the highest number of comments. The next three scores were all students (#27 – $C_b = 54.8$, #23 – $C_b = 25.3$, and #18 – $C_b = 23.4$) who took an active role and served as the go-between in discussions between teachers and scientists.

When these results are combined with the outcomes of the *core-periphery* analysis and the user rankings, the larger picture of the characteristics of principal drivers in the network emerges. While Student B (#19) may not have been the most “chatty,” making seven comments and receiving nine, the location of these posts occurred at critical junctures within the larger discussion space and engaged some of the most active users, including Scientist A (#5), Teacher B (#15), and Student C (#21). Student B (#19) also frequently served as a go-between for two different user groups (teachers and scientists). Both of these

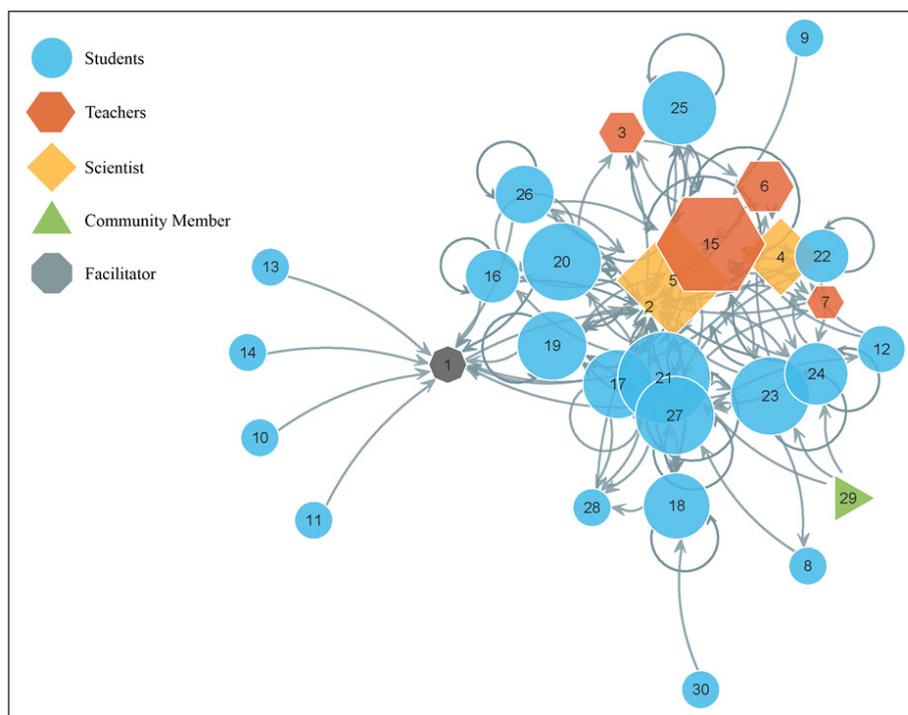


Figure 3: King Tide Investigation Network. Each symbol represents one individual and size is scaled by the number of postings that individual made. Shapes represent user types.

results contribute to Student B’s higher k-index ranking (k_s index score = 14). The results are similar for Student D (#27), who also ranked highly (first quartile, $k_s = 14$) with modest levels of commenting ($n = 11$) and frequent brokerage engagement with other user types.

Looking at user attributes across quartiles based on k-index scores (range: 0–23), each of the top three contains a mix of both adult and student users, providing evidence to support the non-hierarchical nature of the project’s design. The top quartile includes Teacher B, Scientists A and B, and a staff member from the lead organization, each of whom have an equally high k-index score, indicating that they are the innermost core of the network. Four students from School 1 round out the top quartile (k_s range: 14–23), including Student C.

When quartile membership was compared with the discourse coding, a clear pattern emerged. The number of posts was highest for those in the first quartile (i.e., those with the highest k-index scores), followed by those in the second, followed by those in the third, and then those in the fourth (Figure 4). This pattern was consistent for the amount of discourse coded overall, as well as for each knowledge-building construct. In each case, there was a precipitous drop between the amount of discourse coded for the first and remaining quartiles, with only a few comments posted by members in the second through fourth quartiles. This pattern was also reflected in the number of *questions* posed, the *knowledge connections* made, and the amount of *conceptual advancement*. Users in the first quartile made the greatest number of posts containing categories of both higher-level discourse (e.g., full explanation, strong advancement, perpetuation) as well as those that did not contribute to meaningful knowledge-building (e.g., no knowledge, no advancement, beyond scope).

To learn more about how individual users functioned to drive the investigation, a final set of analyses were conducted to focus on individuals in the upper quartile. This revealed that half of the posts across the entire investigation were made by Scientist A and Teacher B, who contributed 25% and 24% of the posts to the investigation, respectively. Given the prevalence and quality of posts by these two members, it is not surprising that they were

among the highest k-index rankings. As such, the differential pattern of results between the first and remaining quartiles in Figure 4 is due, in part, to these two individuals. Though the difference between the first quartile and others is dampened if these users are removed, the overall pattern remains.

Reviewing the individual k-index scores of others in the first quartile revealed a second possible profile for those who drive an investigation. Scientist B had a k-index score equal to that of Scientist A and Teacher B, though the number of comments submitted by Scientist B was far lower. Looking at the remaining members of the first quartile revealed that Student B also had a lower number of posts, relative to others in the quartile, but a high k-index score. To try to understand more about how these members accomplished such high scores, we compared the overall profile of comments used by these low frequency members and other members from the first quartile. Two differences were noted. The two low frequency members posted a balanced number of initiations and responses during the investigation, while all other first quartile members posted many more responses than initiations. Comments from low frequency members were also the most successful at generating a response from other members of the investigation. There were no notable differences in how low frequency members used specific types of discourse.

Conclusions

This study explored an investigation from the *WeatherBlur* online citizen science community to determine the kinds of interactions that were most productive in building knowledge online, and to identify which individuals served as the key drivers for the investigation. Similar patterns of results were found overall in the amount and type of content provided by students, teachers, and scientists. Each initiated conversation at a similar rate, and scientists and students provided a similar portion of responses throughout. Broadly speaking, the use of comments within each knowledge-building construct was also similar across groups. These findings provide evidence to support the successful implementation of the project’s non-hierarchical design and replicate the results

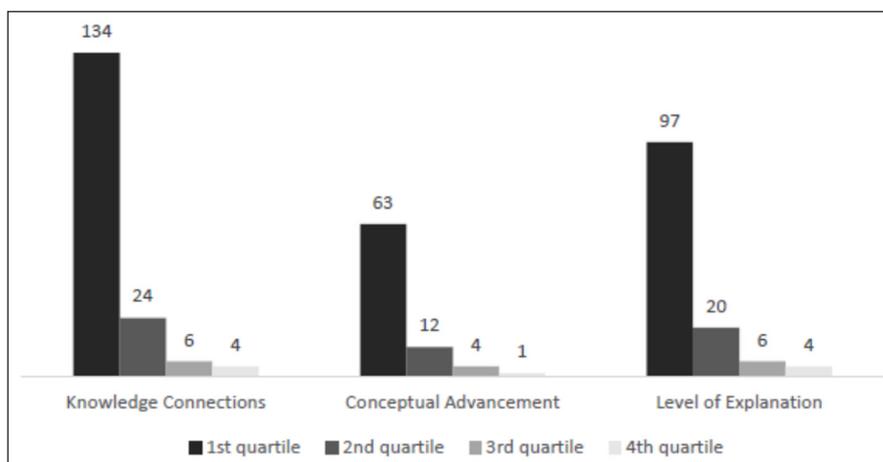


Figure 4: Number of responses by quartile and knowledge-building construct.

of prior work on the *WeatherBlur* project (Kermish-Allen, Peterman, and Bevc 2018).

The results from this study provide potential design recommendations for those interested in encouraging knowledge-building within online citizen science communities. For example, successful strategies for encouraging knowledge-building included conversation about *analysis*, the use of *widening, then deepening* comments, and providing *some advancement*. *Deepening* comments were those used most often to respond to conversation, and thus a successful strategy for sustaining online discourse. Online citizen science projects that are designed to foster any of these kinds of interactions have the potential to catalyze knowledge-building.

Questions were not found to encourage knowledge-building, given that they occurred infrequently and that most questions remained unanswered. Research suggests that questions do not function to elicit the kinds of explanations likely to help build knowledge until the high school years (Lai and Law 2013). Our data replicate these findings in the context of online citizen science and suggest that a deliberate and age-based approach should be used when determining whether and how to encourage the use of questions to attempt to build knowledge in online citizen science projects. Questions have been found to serve a meaningful purpose in knowledge-building among older students and adults (Lai and Law 2013), and as such additional work is needed to determine how questions function in the context of citizen science communities comprising these audiences.

Most conversations were initiated during the beginning and end stages of the investigation process, indicating that these stages of a citizen science investigation might hold the greatest potential for knowledge-building. This result is similar to that found in a recent study of online problem-solving behaviors, in which facilitation was found to be critical in the beginning and ending stages of the process (Huang et al. 2018). This study is limited in the aggregated analysis of user interactions. Future research may benefit from a closer examination of differences across these investigation stages.

Utilizing similar strategies to document the evolution of knowledge-building within an online community over longer periods of time is another opportunity for further research. The current analysis took place within the context of an investigation that lasted seven weeks. Other online citizen science projects include volunteers who participate for months or years. Applying this approach to the interactions of those who participate in longer-term projects has the potential to contribute new understanding about how knowledge-building and the roles of individuals who serve as drivers for citizen science investigations change over time.

Results from both the discourse and social network analysis identified principal drivers within the investigation, indicating that these individuals were not exclusively those who posted at high frequencies. Drivers were also those in key positions to connect users within the discussion space, linking users from different stakeholder groups. When looking closely at the connections and

roles across the network, students demonstrated a broad variety of brokerage potential (*betweenness centrality*), including positions among fellow students and fulfilling key roles in bridging conversations with teachers and scientists. The data from this study also reinforce a common trend in online citizen science projects whereby a minority of online participants produce a majority of the work (Aristeidou et al. 2015; Curtis 2018; and Luczak-Roesch et al. 2014). Even so, the current study is potentially limited as a result of two super users who contributed half of the online discourse, given that results might be specific to these individuals. Future work to replicate the patterns in the current study are needed. As this type of analysis evolves, the combination of discourse and network analysis can be used to reveal and support the more nuanced dimensions of these interactions, particularly those that might otherwise be overlooked if these methods were used independently.

While the discourse and social network analyses used different data and coding schemes to quantify participation in the investigation, the triangulation of results from these analyses provides confidence in our findings. Specifically, those who used traditional knowledge-building language to advance conversation were also those who were the drivers of the investigation based on the network analysis. The network structure for *WeatherBlur* was distributed and mirrors the opportunistic structure identified as an optimal knowledge-building context (Zhang et al. 2009). As such, this study offers a contribution to the knowledge-building literature overall and provides insight into the ways that knowledge-building functions in the context of citizen science. To our knowledge, this is the first study to use knowledge-building constructs to better understand the underlying connections across an online citizen science community. Applying knowledge-building theory to online citizen science projects seems intuitive, particularly for projects that engage youth as volunteers. This study provides initial evidence to demonstrate the potential for youth, teachers, and scientists to build knowledge together in citizen science investigations. The fact that youth serve as drivers reiterates the results found in knowledge-building classrooms and extends those findings through both the addition of scientists to the online community and the focus on citizen science.

Acknowledgements

We would like to thank the following researchers for their help in coding online discourse: Katherine Ciccarelli, Jen Gathings, Maren Harris, and Noelle Wyman Roth. This work was funded by the National Science Foundation, DRL # 1530465.

Competing Interests

Karen Peterman and Christine Bevc do not have competing interests to state. Ruth Kermish-Allen is the Principal Investigator of the *WeatherBlur* program, which provided the context for this study. However, her contributions to the study did not extend to data coding and analysis, thus the authors do not consider her interests to have influenced the research results.

Author Contributions

Karen Peterman served as the lead author on this manuscript and the lead analyst on the discourse analysis. Christine Bevc served as the lead analyst and author for the social network analysis. Ruth Kermish-Allen contributed to introductory text and provided guidance on the framing of the article overall to ensure that it depicted the *WeatherBlur* project accurately.

References

- Aristeidou, M, Scanlon, E and Sharples, M.** 2015. Weather-it missions: A social network analysis perspective of an online citizen inquiry community. In: *Design for teaching and learning in a networked world*, 3–16. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-24258-3_1
- Aristeidou, M, Scanlon, E and Sharples, M.** 2017. Profiles of engagement in online communities of citizen science participation. *Computers in Human Behavior*, 74: 246–256. DOI: <https://doi.org/10.1016/j.chb.2017.04.044>
- Bereiter, C and Scardamalia, M.** 2003. Learning to work creatively with knowledge. *Powerful learning environments: Unravelling basic components and dimensions*, 55–68.
- Bereiter, C and Scardamalia, M.** 2016. “Good Moves” in knowledge-creating dialogue. *Qwerty-Open and Interdisciplinary Journal of Technology, Culture and Education*, 11(2): 12–26.
- Butts, CT.** 2008. Social network analysis with sna. *Journal of Statistical Software*, 24(6): 1–51. DOI: <https://doi.org/10.18637/jss.v024.i06>
- Chen, B, Scardamalia, M and Bereiter, C.** 2015. Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning*, 10(4): 345–366. DOI: <https://doi.org/10.1007/s11412-015-9225-z>
- Connors, JP, Lei, S and Kelly, M.** 2012. Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Annals of the Association of American Geographers*, 102(6): 1267–1289. DOI: <https://doi.org/10.1080/00045608.2011.627058>
- Crall, AM, Newman, GJ, Jarnevich, CS, Stohlgren, TJ, Waller, DM and Graham, J.** 2010. Improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions*, 12(10): 3419–3428. DOI: <https://doi.org/10.1007/s10530-010-9740-9>
- Curtis, V.** 2015. *Online citizen science projects: An exploration of motivation, contribution and participation* (Doctoral dissertation, The Open University).
- Curtis, V.** 2018. Patterns of participation and motivation in Folding@home: The contribution of hardware enthusiasts and overclockers. *Citizen Science: Theory and Practice*, 3(1): 5, 1–14. DOI: <https://doi.org/10.5334/cstp.109>
- Freeman, LC.** 1978. Centrality in social networks conceptual clarification. *Social Networks*, 1(3): 215–239. DOI: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Hakkarainen, K.** 2003. Emergence of progressive-inquiry culture in computer-supported collaborative learning. *Learning Environments Research*, 6(2): 199–220. DOI: <https://doi.org/10.1023/A:1024995120180>
- Hakkarainen, K and Sintonen, M.** 2002. The interrogative model of inquiry and computer-supported learning. *Science & Education*, 11(1): 25–43. DOI: <https://doi.org/10.1023/A:1013076706416>
- Handcock, MS, Hunter, DR, Butts, CT, Goodreau, SM and Morris, M.** 2008. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1): 1548. DOI: <https://doi.org/10.18637/jss.v024.i01>
- Hmelo-Silver, CE and Barrows, HS.** 2008. Facilitating collaborative knowledge building. *Cognition and instruction*, 26(1): 48–94. DOI: <https://doi.org/10.1080/07370000701798495>
- Huang, HC, Cheng, TCE, Huang, WF and Teng, CI.** 2018. Who are likely to build strong online social networks? The perspectives of relational cohesion theory and personality theory. *Computers in Human Behavior*, 82: 111–123. DOI: <https://doi.org/10.1016/j.chb.2018.01.004>
- Kermish-Allen, R, Peterman, K and Bevc, C.** 2018. The utility of citizen science projects in K-5 schools: Measures of community engagement and student impacts. *Cultural Studies of Science Education*. DOI: <https://doi.org/10.1007/s11422-017-9830-4>
- Lai, M and Law, N.** 2013. Questioning and the quality of knowledge constructed in a CSCL context: A study on two grade-levels of students. *Instructional Science*, 41(3): 597–620. DOI: <https://doi.org/10.1007/s11251-012-9246-1>
- Liberatore, A, Bowkett, E, MacLeod, CJ, Spurr, E and Longnecker, N.** 2018. Social Media as a Platform for a Citizen Science Community of Practice. *Citizen Science: Theory and Practice*, 3(1). DOI: <https://doi.org/10.5334/cstp.108>
- Lin, F and Chan, CK.** 2018. Examining the role of computer-supported knowledge-building discourse in epistemic and conceptual understanding. *Journal of Computer Assisted Learning*. DOI: <https://doi.org/10.1111/jcal.12261>
- Luczak-Roesch, M, Tinati, R, Simperl, E, Van Kleek, M, Shadbolt, N and Simpson, RJ.** 2014, June. Why Won't Aliens Talk to Us? Content and Community Dynamics in Online Citizen Science. In: *ICWSM*.
- Newman, G, Crall, A, Laituri, M, Graham, J, Stohlgren, T, Moore, JC, Kodrich, K and Holfelder, KA.** 2010. Teaching citizen science skills online: Implications for invasive species training programs. *Applied Environmental Education and Communication*, 9(4): 276–286. DOI: <https://doi.org/10.1080/1533015X.2010.530896>
- Ono, E, Ikkati, Y and Enoto, T.** 2018. Increasing crowd science projects in Japan: Case study of online citizen participation. *International Journal of Institutional Research and Management*, 2(1): 19–34.
- Oshima, J, Oshima, R and Matsuzawa, Y.** 2012. Knowledge building discourse explorer: A social network

- analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60(5): 903–921. DOI: <https://doi.org/10.1007/s11423-012-9265-2>
- Oshima, J, Scardamalia, M and Bereiter, C.** 1996. Collaborative learning processes associated with high and low conceptual progress. *Instructional Science*, 24(2): 125–155. DOI: <https://doi.org/10.1007/BF00120486>
- Palonen, T and Hakkarainen, K.** 2013, April. Patterns of interaction in computer supported learning: A social network analysis. In: *Fourth international conference of the learning sciences*, 334–339.
- R Core Team.** 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Resendes, M, Scardamalia, M, Bereiter, C, Chen, B and Halewood, C.** 2015. Group-level formative feedback and metadiscourse. *International Journal of Computer-Supported Collaborative Learning*, 10(3): 309–336. DOI: <https://doi.org/10.1007/s11412-015-9219-x>
- Scardamalia, M and Bereiter, C.** 1994. Computer support for knowledge-building communities. *The journal of the learning sciences*, 3(3): 265–283. DOI: https://doi.org/10.1207/s15327809jls0303_3
- Sun, Y, Zhang, J and Scardamalia, M.** 2010. Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science*, 38(2): 147–171. DOI: <https://doi.org/10.1007/s11251-008-9082-5>
- Tinati, R, Luczak-Roesch, M, Simperl, E and Hall, W.** 2016, May. Because science is awesome: Studying participation in a citizen science game. In: *Proceedings of the 8th ACM Conference on Web Science*, 45–54. ACM. DOI: <https://doi.org/10.1145/2908131.2908151>
- Tinati, R, Simperl, E, Luczak-Roesch, M, Van Kleek, M and Shadbolt, N.** 2014. Collective Intelligence in Citizen Science—A Study of Performers and Talkers. *arXiv preprint arXiv:1406.7551*.
- Tinati, R, Van Kleek, M, Simperl, E, Luczak-Rösch, M, Simpson, R and Shadbolt, N.** 2015, April. Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 4069–4078. ACM. DOI: <https://doi.org/10.1145/2702123.2702420>
- Wenger, E.** 2000. Communities of practice and social learning systems. *Organization*, 7(2): 225–246. DOI: <https://doi.org/10.1177/135050840072002>
- Zhang, J, Scardamalia, M, Reeve, R and Messina, R.** 2009. Designs for collective cognitive responsibility in knowledge-building communities. *Journal of the Learning Sciences*, 18(1): 7–44. DOI: <https://doi.org/10.1080/10508400802581676>

How to cite this article: Peterman, K, Bevc, C and Kermish-Allen, R. 2019. Turning the King Tide: Understanding Dialogue and Principal Drivers in an Online Co-Created Investigation. *Citizen Science: Theory and Practice*, 4(1): 3, pp. 1–12. DOI: <https://doi.org/10.5334/cstp.189>

Submitted: 09 July 2018

Accepted: 26 November 2018

Published: 13 February 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

] *Citizen Science: Theory and Practice* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 