



# Characterizing Student-Driven Research Investigations Contributed to the GLOBE Program Citizen Science Initiative in a Formal Education Context

ANN MARTIN

KATHERINE MILLER-BAINS

JULIE MALMBERG

LIN CHAMBERS

KEVIN CZAJKOWSKI

*\*Author affiliations can be found in the back matter of this article*

## CASE STUDIES

]u[ubiquity press

## ABSTRACT

The Global Learning and Observations to Benefit the Environment (GLOBE) Program offers citizen science opportunities to participants of all ages, with a focus on youth in formal classroom contexts. This study uses student investigation research reports and posters submitted to the 2018 International Virtual Science Symposium (IVSS) and Student Research Symposium (SRS) as testbeds for characterizing student-driven Earth system citizen science investigations. Secondly, this study aimed to capture GLOBE's alignment to existing citizen science outcomes frameworks in the literature, which have primarily focused on adults and non-formal settings. Based on a literature review, the evaluation team identified 89 potential characteristics in 27 categories to typify investigations from both formal education and citizen science perspectives. We coded the artifacts from 207 student projects, conducted quantitative analysis of frequencies, and performed a semantic network analysis. By using this networking approach, we conceptually mapped several clusters of co-occurring characteristics, defining a descriptive framework for GLOBE projects. We identified three tiers of citizen science projects, increasing in the sophistication of participants' demonstrated science practices. The framework includes additional components that reflect student citizen scientists' thoughtfulness and connection to context as well as their projects' reflection of their motivation and self-efficacy. Through these findings, we have identified areas where student citizen scientists would benefit from further support, and suggest here further research to incorporate the experiences of students into the broader understanding of citizen science outcomes.

## CORRESPONDING AUTHOR:

**Ann Martin**

Oak Ridge Associated  
Universities, US

[ann.martin@orau.org](mailto:ann.martin@orau.org)

## KEYWORDS:

network analysis; semantic analysis; characteristics; student investigations; formal education

## TO CITE THIS ARTICLE:

Martin, A, Miller-Bains, K, Malmberg, J, Chambers, L and Czajkowski, K. 2022. Characterizing Student-Driven Research Investigations Contributed to the GLOBE Program Citizen Science Initiative in a Formal Education Context. *Citizen Science: Theory and Practice*, 7(1): 11, pp. 1–16. DOI: <https://doi.org/10.5334/cstp.480>

## INTRODUCTION

The Global Learning and Observations to Benefit the Environment (GLOBE) Program, which began in 1995, is an international science and education program that encourages students and the public to collect and share data about the Earth System (<https://www.globe.gov>; Finarelli 1998). GLOBE is sponsored by the National Aeronautics and Space Administration (NASA); supported by the National Science Foundation (NSF), the National Oceanic and Atmospheric Administration (NOAA), and the US Department of State (DOS); and implemented by the University Corporation for Atmospheric Research (UCAR) through a cooperative agreement with NASA (Grant # 80NSSC19M0120). Internationally, GLOBE is implemented through government-to-government agreements with each partner country. GLOBE originally began as a school-based citizen science program, but expanded to all interested citizen scientists in participating countries in 2016 through the GLOBE Observer app. Now in 126 countries, more than 42,000 teachers and 214,000 citizen scientists have participated in GLOBE.

GLOBE therefore sits at the interface of citizen science and formal education, and presents an opportunity to examine the potential for formal learning outcomes from participation in citizen science. GLOBE students or classrooms drive their own investigations, with assistance from teachers, while participating in structured aspects echoing the Bonney et al. (2009) framework's "collaborative" and "contributory" activities. For instance, GLOBE's so-called campaigns invite students to participate in bursts of citizen science data collection activities led by scientists and educators, and campaigns provide prompts, activities, and community supports.

## BACKGROUND AND CONTEXT

The study described in this paper draws from two major GLOBE initiatives, the International Virtual Science Symposium (IVSS) and the Student Research Symposia (SRS). Both initiatives provide a structured opportunity for students to present research investigations and participate in the communal aspects of science. This study's dataset comprises student entries into the 2018 IVSS and SRS.

The GLOBE IVSS began in 2012 as part of an NSF Innovative Technology Experiences for Students and Teachers (ITEST) award (Grant No. 0929725). A virtual platform was developed on the GLOBE website for students to upload their projects and comment on other student projects (Malmberg and Maull 2013). In 2013, submissions were opened to schools from all GLOBE countries. In 2016,

the IVSS was re-launched with the addition of virtual badges based upon demonstration of key science practices (e.g., making an impact).

The US Regional GLOBE Student Research Symposia (SRS), by contrast, are in-person events for students within the United States. Annual SRS events began in 2015 with funding from NSF (Grant #1546713) and are currently funded by a grant from NASA (Grant #80NSSC18K0135) and Youth Learning As Citizen Environmental Scientists (YLACES), a granting organization associated with GLOBE (<https://www.ylaces.org/>). Annual in-person events were held at six regional locations around the country from 2016 to 2019 (2020–2022 SRS were cancelled due to COVID-19, although funding will be provided for small in-person events in 2022). At these events, GLOBE student investigators meet students from other states, discuss their research with scientists, and learn about STEM careers.

## STUDENT LEARNING OUTCOMES IN CITIZEN SCIENCE AND AUTHENTIC SCIENCE INVESTIGATIONS

As noted above, the citizen science research community has explored a variety of learning and attitudinal outcomes, while GLOBE also considers the formal learning perspective. As a preliminary step to measuring learning outcomes, in this study we seek to explore and identify a potential range of them.

Bonney et al.'s (2009) model identified potential outcomes and indicators in several categories, including awareness/knowledge/understanding, engagement or interest, skills, attitudes, and behaviors. Phillips et al. (2018) extended this argument into a framework that translates outcomes developed from the informal science learning literature into a range of citizen science-specific learning outcomes. These include i) interest in science and the environment; ii) self-efficacy for science and the environment; iii) motivation for science and the environment; iv) knowledge of the nature of science; v) skills of science inquiry; and vi) behavior and stewardship. GLOBE learning outcomes may also depend on students' interest and motivation, which influence the topic they choose to explore. Topic selection is also influenced by established GLOBE protocols and activities. As in Chase and Levine's (2016) framework, student motivation and topic selection may be driven by biophysical or environmental characteristics, student awareness of opportunities to pursue the topic within GLOBE, current or former monitoring, and sociocultural aspects such as the opportunity to participate in a broader GLOBE campaign with other youth.

Much of the literature has focused on adult and informal learning within citizen science (e.g., Aristeidou and Herodotou 2020; Bonney et. al. 2009; Chase and

Levine 2016; National Academies 2018; Phillips et al. 2018), or student learning from participation in science fairs or inquiry instruction rather than student learning from a citizen science investigational context (e.g., DeLisi et al. 2020; Houseal, Abd-El-Khalick, and Destafano 2014; Craven and Hogan 2008). Aristeidou and Herodotou (2020) conducted a systematic review of the effects of citizen science on learning and literacy outcomes, and noted a lack of studies of citizen science in formal settings. This study begins to bridge that gap and to consider what features of citizen science investigations may be aligned with learning outcomes.

A National Academies report on *Learning Through Citizen Science* (2018) proposes a model in which citizen science provides experiential opportunities aligned to what we already know about human learning: engagement with other people, constructing knowledge through activities, developing appropriate content knowledge, interacting with data, and using both the tools and the practices of science and inquiry. They further incorporate motivation and interest (Geoghegan et al. 2016; Frensey et al. 2017) as key contextual factors that mediate possibilities for learning. This report also mentions interest in the potential for citizen science to develop disciplinary knowledge but a lack of evidence for such outcomes. A synthesis paper by Phillips et al. (2019) similarly observes a gap in the field surrounding the goal of many citizen science projects to develop understanding of the nature of science, with only weak evidence for achieving this.

Inquiry instruction and student-centered authentic science activities within formal settings provide the best analog to the GLOBE citizen science program. DeLisi et al. (2020) find that student learning as a result of science fair participation is most associated with opportunities to participate in evaluating and critiquing findings of projects. These authors argue that science fairs, like GLOBE or other citizen science opportunities accessible to youth, are an unusual opportunity for students to take ownership over a complete scientific investigation and to think critically about their own investigative work. Houseal, Abd-El-Khalick, and Destefano (2014) established student learning gains in a formal teacher-student-scientist partnership initiative and found that giving students agency in their research investigations and providing true authenticity can impact learning.

This echoes the broader perspective of Minner, Levy, and Century (2010), who conducted a research synthesis of K-12 inquiry instruction across studies from 1984 to 2002, and found that emphasizing active thinking, assigning students responsibility for learning, or drawing conclusions from data can positively influence student learning. Craven and Hogan (2008) summarize experience with science fairs

similarly to the citizen science community's findings on citizen science outcomes: The nature of investigations, and the extent to which learners are engaged in real-life issues at the nexus of science content and personal/societal motivations, differentiate high-quality experiences from rote ones.

## PURPOSE AND RESEARCH QUESTIONS

As noted above, GLOBE is situated in a unique space within citizen science. While GLOBE is scientifically rigorous, with protocols for data collection by citizen scientists and a goal of large-scale environmental data collection, SRS and IVSS are also situated in formal education contexts. Investigations are driven by students, as individuals, as a class, through assignments by their teachers, or via participation in organized campaigns.

For this study, we were interested in asking key questions within this specific citizen science/formal learning space: What are the characteristics of student investigations that are supported by The GLOBE Program? How can we understand student engagement with content and the scientific process by applying knowledge about citizen science frameworks to this educational, student-driven context?

This paper describes a mixed-methods empirical study of GLOBE student investigations from the 2018 IVSS and SRS opportunities. The IVSS and SRS initiatives are ideal for this study because a large number of students participate in a structured way, and each initiative collects standard artifacts. Our goal was to identify and explore the characteristics of student research investigations submitted to GLOBE. More specifically, we intended to situate those characteristics within the broader landscape and scholarship of citizen science. We sought to identify ways that the characteristics of student investigations may echo the traits of traditional citizen science programs that may involve adults or professional science projects more explicitly. Ultimately, our goal was to create a descriptive framework of citizen science characteristics suitable for contexts like GLOBE's education-driven, student-led approach to citizen science.

## METHODS AND DATA ANALYSIS

The data sources for this study included the student reports and posters submitted to the 2018 IVSS online events and SRS in-person events. IVSS projects were posted on the GLOBE program website (<https://www.globe.gov/news-events/globe-events/virtual-conferences/2018-international-virtual-science-symposium>). SRS projects were presented in person,

and hosts at each event took high-resolution photographs of the poster presentations. IVSS data was gathered from the GLOBE website by the research team at ORAU, with metadata provided by author Malmberg. Photographs of SRS posters were provided by Haley Wicklein and Jen Bourgeault (University of New Hampshire/GLOBE US Country Coordinator Office).

As presented in the following sub-sections, the study mixes standard quantitative and qualitative methods with semantic network analysis techniques, resulting in a richly informed qualitative framework for understanding GLOBE student investigations. First, the researchers conducted a literature review to identify relevant frameworks and to develop a list of codes and characteristics relevant to GLOBE projects. The code list was refined, and each project in the dataset was qualitatively coded for the presence or absence of each characteristic. Quantitative analysis of the frequency of codes across projects provided insights into the most- and least-common approaches in GLOBE student investigations, and semantic network analysis of co-occurring codes identified clusters of characteristics indicative of a typology of student investigations. Final theoretical coding of the networked clusters resulted in a theory-driven descriptive framework. The following sub-sections elaborate on each of these steps in greater detail.

### CHARACTERISTICS OF THE DATA

113 entries into the IVSS and 110 entries into the SRS events were considered for inclusion. Ultimately, after excluding projects lacking a report or poster and those projects that were illegible (such as blurry poster photographs), a total of 207 projects were carefully examined. Of these, 97 (47%) were drawn from IVSS, 96 (46%) from SRS, and an additional 14 (7%) had been submitted to both opportunities.

Age/grade band information was available for only 203 of the projects ([Table 1](#)), using GLOBE's international definitions aligned to approximate age groups. Because

only one undergraduate project was submitted, that group has been removed from most analyses. In part because SRS events took place within the United States, the majority of projects were completed by US students.

### ETHICS AND CONSENT

The study and all protocols were reviewed by the Oak Ridge Site-wide Institutional Review Board and were determined not to constitute human subjects research because of the exclusive use of archival data. The data and metadata used for analyzing the IVSS reports were publicly available on the GLOBE website. There was no interaction with the students. For the SRS reports, photographs of student research posters were taken by members of the GLOBE team at in-person SRS events under a media release signed by parents/guardians.

In both cases, the project materials do contain personally identifying information. Materials were stored on an encrypted system at ORAU. Beyond IRB requirements, the researchers were committed to ensuring that no students would recognize their own projects in the research, including this manuscript.

### LITERATURE REVIEW TO IDENTIFY RELEVANT FRAMEWORKS FOR CODING STUDENT INVESTIGATIONS

The study began with a thorough review of relevant literature in citizen science and in public participation in scientific research. From an initial list of prominent papers, the authors followed citations to approximately 20 reports and journal articles. The authors created an annotated bibliography and an initial list of codes, representing GLOBE-relevant characteristics of participant investigations, motivations, experiences, and outcomes in citizen science ([Table 2](#)). The codes on this initial list were drawn from the various frameworks and as a result in many cases overlapped. The authors identified the most

BY AGE/GRADE BAND	N (OUT OF 203)	%
Lower primary (grades k–2, ages 5–8)	5	2%
Upper primary (grades 3–5, ages 8–11)	9	4%
Middle school (grades 6–8, ages 11–14)	82	40%
Secondary school (grades 9–12, ages 14–18)	106	52%
Undergraduate	1	<1%
BY COUNTRY	N (OUT OF 207)	%
United States	128	62%
Outside United States (IVSS only)	79	38%

**Table 1** Grade band and country of students submitting projects.

SOURCE	EXAMPLE CATEGORIES
Bonney et al. 2009	project type; defining questions; interpreting data; developing explanations; disseminating conclusions
Chase and Levine 2016	biophysical and geographical factors; geographical scale; temporal scale; group self-organization; protocol training; collection methods; social factors
Edelson et al. 2013	designing solutions; communicating information
Freitag et al. 2016	planning phase; prior training; assistance from professional; validation; cross-comparison
Kastens 2014a	unstructured observation; student-collected data sets; well-structured problems
Kastens 2014b	direct knowledge; mediated knowledge; creation of representation; interpretation of representation
NASEM 2018	ground truthing; action project; education project; scientific practices
NACEPT 2016	community engagement; filling gaps in data sets; scale
Phillips et al. 2018	interest; self-efficacy; inquiry skills; stewardship behavior; community action
Shirk et al. 2012	degree of participation
Tweddle et al. 2012	analysis and reporting; share data; take action in response to data; evaluate/reflect
The GLOBE Program 2018a,b	spatial scale; GLOBE protocols; context and relevance; connecting to a STEM professional; interscholastic connections; engineering solutions
Wiggins and Crowston 2011	action/conservation; data validity

**Table 2** Literature sources for GLOBE investigation codes.

useful constructs for characterizing and organizing citizen science projects within the GLOBE context.

Closely related or overlapping codes were combined, and the entire set was organized loosely into related dimensions and themes. Throughout the code refinement process, labels were used to indicate the initial source of the code, which remained traceable throughout the study; these source labels remain attached to each code listed in Supplemental File 1: Frequency of Individual Code Occurrences by Project, by Grade Band, and by Country.

### DEVELOPMENT AND REFINEMENT OF CODES

The main coder (author Martin) started with the initial list of codes pulled from the literature, already organized into top-level dimensions and secondary themes, and reviewed them, which resulted in a few changes. The coder tested this initial list on a randomly-selected subset of 10 IVSS and 10 SRS projects, revised the list, and sent the codes for final review by authors Malmberg, Chambers, and Czajkowski. This final list of codes was then applied to a new random sample of 10 IVSS and 10 SRS projects for final testing before moving forward. The final list of codes, organized by dimension and theme, are included (along with findings) in Supplemental File 1.

Two types of codes require special discussion. The first category are codes and characteristics that are specific to the GLOBE context. Because some students chose to submit a report or poster that was not completed within the GLOBE program, codes were created to identify whether each

project was a GLOBE project. Projects were also assigned to GLOBE “spheres” or topic areas (hydrosphere, biosphere, atmosphere, and soil), and tagged with the type of technology used to collect the data, because this is of special interest to GLOBE. Finally, we added codes related to GLOBE’s virtual badges, indicating project elements like interaction with a scientist or GLOBE students at another school.

Another category of codes, which were of particular interest to the authors, are drawn from Kastens (2014a,b). These brief concept papers outline a framework describing youths’ progression as they develop the skills and habits of mind of scientists who work with data. We adapted Kastens’ framework to differentiate between student projects that display non-mediated knowledge (data as collected by students in raw format, or represented in simple tables), simply mediated knowledge (data translated to a representation or model), and more sophisticated demonstrations of mediation of knowledge (data translated to a representation or model with benchmarks or other interpretive elements to guide understanding).

### CODING THE FULL DATASET AND INTERRATER RELIABILITY

Once the codes were fully validated and finalized, the coder then applied them to the entire set of projects, including re-coding the randomly-selected projects in the testing phase. Each project was carefully examined and coded by a single evaluator who assessed whether each code was or was not present. The evaluator sought specific evidence for the



characteristic in the artifacts. Thus, the question answered by coding was never whether the students completed a given task or learned a particular skill, but whether their research artifact provided a demonstration or evidence of a given facet. The absence of evidence in these artifacts is not evidence of absence in the students' experiences or their learning. While this is a potential weakness of the study, it is also reflective of the SRS and IVSS, in which artifacts are incomplete representations of the citizen science experience.

Notably, some codes are independent of others and could be selected regardless of other characteristics, such as the codes related to the topic. As many such codes as were applicable were selected. In other cases, though, the coder would likely only select one from a set of codes. For example, a given investigation would be coded as either a planned investigation or an unplanned/open-ended investigation, but not as both. In other areas, the coder strived to select only the most salient option from a set of codes, such as the motivation for the selection of a topic.

To investigate the reliability of coding, a subset of 47 projects were double coded by an independent coder (author Miller-Bains). Unlike the main coder, the independent coder had no prior exposure to GLOBE, and required some training to recognize key elements. Following this interrater reliability sub-study, several adjustments were made to the code list and relevant projects were re-coded.

Based on the results of two phases of reliability coding, estimates of rater agreement were calculated using R Core Team (2021). Details of the interrater reliability study are reported in Supplemental File 2: Tabulation of Interrater Reliability Scores and Prevalence Indices. In the case of some codes, the agreement values did not reach the lower threshold for acceptable reliability. Overall, agreement varied widely across codes, but some of the variation can be attributed to both the dichotomous nature of the coding scheme and the infrequency of certain codes (Eugenio and Glass 2004). Based on this interrater reliability study, the authors advise caution in the interpretation of the codes, particularly for those who are unfamiliar with the GLOBE program.

## QUANTITATIVE AND NETWORK ANALYSIS

The completed dataset was quantitatively analyzed to study the frequency of each characteristic. These findings were disaggregated by student age/grade band and by country (United States versus Non-United States). The dataset was also analyzed using the tools of social network analysis to identify and cluster characteristics that frequently appeared together (e.g., semantic network analysis; Doerfel 1998) using the open-source Gephi social network analysis software (Bastian, Heymann, and Jacomy 2009).

A visual inspection of the semantic network did not suggest any clear organization (Prell 2012). The Blondel et al.

(2008) method for identifying clusters or subcommunities using the modularity metric was calculated in Gephi to identify groups of co-occurring codes. After reviewing the results of the initial clustering algorithm, some codes were excluded that did not aid in the creation of a framework of student investigation characteristics. For example, the codes related to the topic area (biosphere, soil/pedosphere, hydrosphere, and atmosphere) were independent of the clustering this study sought to examine. Once these codes had been eliminated, the modularity calculation was repeated to identify modularity clusters—sets of codes that often coexisted together in the network, and that were less likely to coexist with codes in any of the other modularity clusters. These clusters, therefore, are indicators of cohesive sets of characteristics of student GLOBE investigations. The evaluator examined each cluster and the codes present within it, revisited the literature sources in [Table 2](#), then conducted so-called second cycle qualitative coding (Saldaña 2015). This phase of theoretical coding identified commonalities among the codes appearing in each cluster, for which the evaluator developed descriptive titles. Theoretical coding further identified similarities and distinctions between each cluster, which suggested ways of grouping and organizing the clusters. The final descriptive framework, presented in the section entitled “Findings,” presents the sub-clusters identified by the algorithm and the larger theory-driven clusters into which these sets were organized by the evaluator.

## CAVEATS AND LIMITATIONS

The most significant limitation to the study is that coding of projects was largely completed by a single rater, albeit one with substantial experience in GLOBE. We attempted to ameliorate this limitation through discussion among the authors, substantial code reviews, and the interrater reliability inquiry. We do suggest caution in drawing conclusions from this limited data. Throughout this process, we noted that knowledge of the GLOBE program was helpful in conducting the study and interpreting the findings. Although our intent is and has been to situate these findings in a context that is of use to the broader citizen science community, GLOBE's specific prominence as a case study is clear.

Second, the Blondel et al. (2008) modularity algorithm used in the Gephi software for the semantic network analysis is nondeterministic. Each iteration results in somewhat different assignments of codes among clusters, and the evaluator has the ability to make some choice in the algorithm's settings, allowing for further variation. The sub-clusters identified and presented in the section entitled “Findings” are based on converging solutions that appeared over multiple runs, but the nondeterministic nature of this technique is a limitation

of the study. Ultimately, the main-level clusters are not based solely on the modularity metrics, but were assigned by the researchers based on careful examination of the quantitative findings, and theory-driven qualitative analyses of the codes within each modularity cluster.

Finally, the student artifacts used were not the ideal data source for some of the characteristics of interest. For instance, we were not able to assess whether students increased their knowledge or skills, but only whether they demonstrated such an increase by discussing or mentioning it in their materials. Thus, this manuscript is most fairly treated as a study of the artifacts resulting from the investigations rather than a full study of the citizen science experience in GLOBE.

## FINDINGS

### QUANTITATIVE PROMINENCE OF CHARACTERISTICS

The full list of 89 codes is presented in Supplemental File 1, along with the frequency of each code across the projects and disaggregated by student grade band and country. While Supplemental File 1 provides a full accounting of all codes used in this project, *Table 3* provides a summarized level of insight into the codes by providing the theme areas and, where applicable, sub-theme areas around which the codes were organized.

In most cases, codes illustrative of more complex scientific thinking and scientific processes became more common as the grade band of the student citizen scientists

THEME AREA	SUB-THEME AREAS (IF APPLICABLE)
Increased engagement/interest	
Increased knowledge	
Increased skill	
Increased student self-efficacy/behavior change	Behavior change Self-efficacy change
Key aspects of the scientific process	Pre-investigation Carrying out Finalizing
More sophisticated aspects of the scientific process	Data limitations Data quality, validation, calibration or investigation Using data and results Broader scientific context Engineering principles
Gathering data	Data collected GLOBE fidelity Deviations from GLOBE Technological collection aids Human senses
Geographic scale	Local Beyond local
Temporal scale	Temporal scale driven by topic Temporal scale driven by length of data collection
GLOBE sphere(s)	
Goals/type of investigation	Exploring a natural system Continuity/gap filling Student-driven No investigation Interdisciplinary

(Contd)

THEME AREA	SUB-THEME AREAS (IF APPLICABLE)
Motivation/context for selection of topic	Relevance to people
	Relevance of people to environment
	Continuation
	Current events
	About GLOBE
	Not a GLOBE project
Broader relevance of subject	
Planned vs. open-ended	Planned
	Unplanned
Student organization	Self-organized
	Joined larger effort
	Team/roles identified
Complexity of hypothesis	Simple hypothesis
	Complex hypothesis
	Weak understanding
Variable control	Controlled variables
	Observed variables
Statistical analysis	Basic stat analysis and interpretation
	No stat analysis or poor analysis/interpretation
	Sophisticated stat analysis/interpretation
Coherence between research question and conclusions	Proper conclusions/successfully addressed
	Over-stretched or erroneous conclusions
Level of structure	Unstructured data
	Structured/systematic data
	Problem with structured scope/range—simple
	Problem that is unstructured/less structured
Mediation of knowledge through representations	Direct knowledge
	Mediated knowledge
	Mis-mediated
Connection from idea to complete investigation	Weak connections
	Reasonable connections
	Low/medium complexity
	High complexity
Careers	
STEM professional relationship	
Interscholastic collaboration	
Considerations of impact/stewardship	Ignored impact
	Local impact
	Actionable but no action
	Local to global
	Next steps
Action taken	

**Table 3** Summary of theme and sub-theme areas of final code list.



increased. For instance, while about 4% of projects overall employed unstructured, exploratory observations as the main learning and investigative strategy, this was more common at the lower and upper elementary levels (20% and 11%, respectively) than among secondary school-aged citizen scientists (3%). In other cases, the opposite was true: While 27% of projects overall discussed actionable information of concern in the local community or environment, this occurred in 60% of lower elementary citizen science investigations. However, because more middle and secondary school students participated overall, these disaggregated findings are only descriptive and suggestive. A study including a larger pool of student investigations would be needed to substantiate differences by grade.

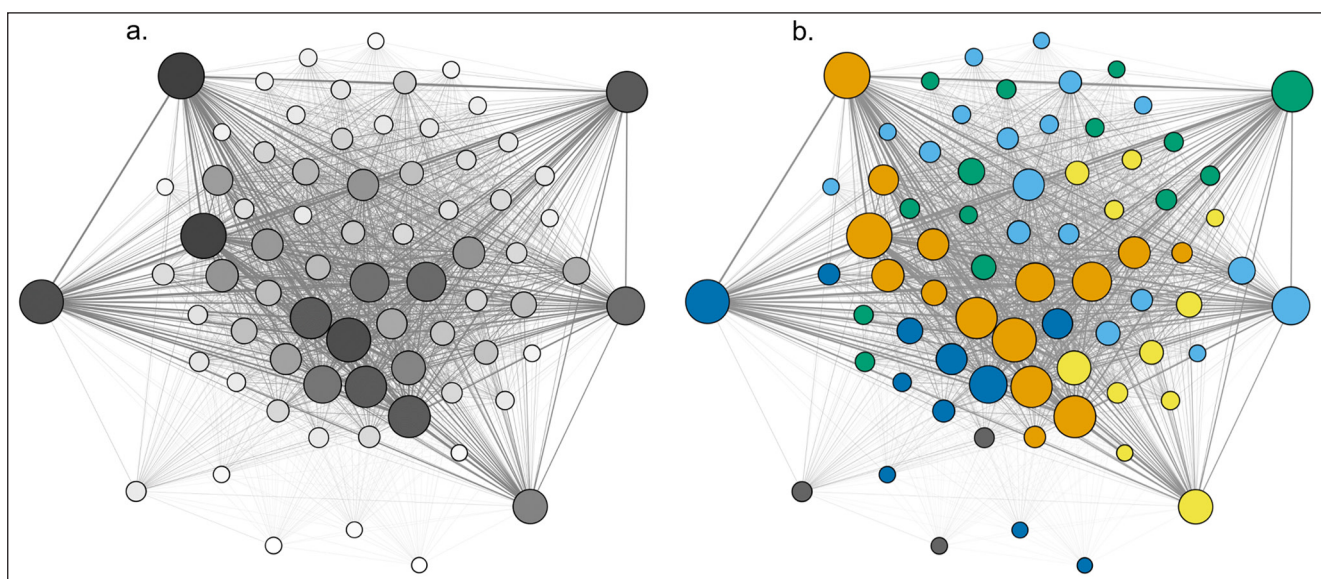
### CLUSTERS AND DESCRIPTIVE FRAMEWORK

The co-occurrences network is displayed in *Figure 1*, which includes only the codes that were retained in the analysis to identify clusters. The left-hand panel indicates the structure of the network before the modularity clustering analysis was applied. This perspective shows each individual code as a node in the network, with radius and color depth indicating the frequency of each code. The number of linkages emanating from each node indicates how commonly that node co-occurs with others, and the thickness of each link indicates how frequently the two nodes connected by that edge co-occurred. This fundamental semantic network data comprised the input used for the next step, the cluster analysis, the result of

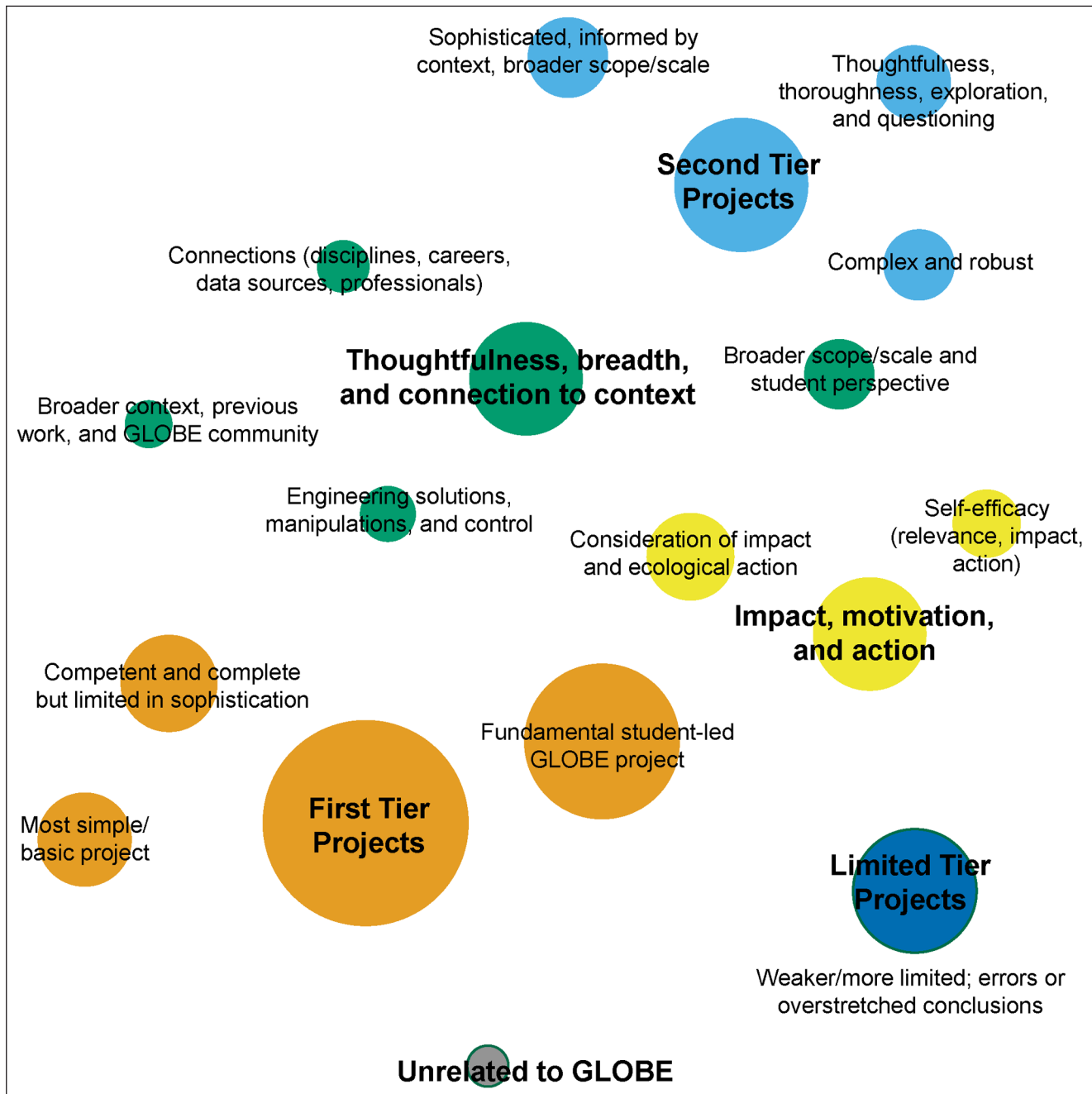
which is displayed in the right-hand panel. The right-hand panel displays this same network with each node in the same location, after the application of the modularity algorithm and the final qualitative coding by the evaluator in the final step of analysis. The color-coding in the right-hand panel indicates the main cluster to which each code was ultimately assigned. These clusters along with subclusters are further described in *Figure 2*.

*Figure 2* illustrates the framework derived from IVSS and SRS projects. It represents each of the clusters and subclusters; the area of each bubble is based on the prominence of that cluster or sub-cluster among the student projects. Note that each project could belong to multiple clusters, and the framework provides a way to describe projects holistically based on multiple elements, rather than a way to assign student citizen science investigations to a specific cluster.

On the basis of a qualitative examination of the types of codes and characteristics in each identified cluster, and on the literature sources in *Table 2*, we identified three tiers of student citizen science investigation, from limited to first to second tier, each increasing in the sophistication of practices and methods captured by the codes within that cluster. The second-tier projects included indicators of more sophisticated, complex, or evaluative thought processes among student investigators. The framework also includes two additional major clusters, which stand apart from the scientific or methodological maturity of the investigation. These additional clusters demonstrate student citizen scientists' thoughtfulness, connection to



**Figure 1** (a) Network mapping of individual codes before cluster analysis, and (b) codes once assigned to clusters based on the network analysis modularity technique and second-cycle theoretical coding of clusters. Color legend is as displayed in *Figure 2*.



**Figure 2** Concept map framework of project typologies based on cluster analysis and theoretical coding.

context, or motivation and self-efficacy to act based on findings. The figure also demonstrates how the clusters related to project content and student focus on impact or action span across those three tiers of projects. The final cluster, displayed at the bottom of *Figure 2*, includes submitted project materials that were unrelated to GLOBE.

The relative prominence of each cluster across the projects is displayed below in *Table 4*. Note once again that each project can fall into multiple clusters. Rather than

being used as a tool to divide projects into categories, this framework provides a way to describe multiple intersecting elements of projects. For instance, a project with serious flaws in the analysis and interpretation of field data can also demonstrate student engagement around environmental action. The cluster frequencies broken down by grade band and by country are provided in Supplemental File 3: Frequency of Project Alignment with Identified Clusters and Sub-Clusters by Grade Band, and by Country.

CLUSTER AND SUB-CLUSTER	N (OUT OF 207)	%
<i>Project unrelated to GLOBE</i>		
Project unrelated to GLOBE	29	14%
<i>Limited-tier projects (limited sophistication)</i>		
Weaker/more limited project with errors, overstretched conclusions, or fundamental weaknesses of design and structure	37	18%
<i>First-tier projects (more sophisticated)</i>		
Demonstrates fundamentals of student-led GLOBE investigations	179	87%
Most simple/basic project	57	28%
Competent and complete project but limited in sophistication	67	32%
<i>Second-tier projects (most sophisticated)</i>		
Complex and robust project	27	13%
More sophisticated project that is informed by context and reflects broader scope/scale and data literacy elements	22	11%
Reflective of student thoughtfulness, thoroughness, exploration, and questioning	9	4%
<i>Additional characteristics: indicators of impact, motivation, and action</i>		
Student consideration of impact and ecological action	64	31%
Student self-efficacy and translation of project into relevance, impact, and action	15	7%
<i>Additional characteristics: indicators of thoughtfulness, breadth, and connections to context</i>		
Investigation involves control/manipulation or engineering solutions	36	17%
Investigation connects to a broader context, previous work, or the larger GLOBE community	10	5%
Investigation demonstrates student connection (disciplines, careers, data sources, context, and STEM professionals)	10	5%
Investigation reflects broader scale/scope and broader student perspective	32	16%

**Table 4** Frequency of project alignment with identified clusters and sub-clusters.

## DISCUSSION AND IMPLICATIONS

### NOTABLE INDIVIDUAL CODES

This section presents a selection of interesting codes and the frequency of their occurrence among the projects, indicating the prominence of selected characteristics among the 2018 investigations (see Supplemental File 1). Since codes are not mutually exclusive, groups of codes in Supplemental File 1 may display totals of less or more than 100%.

Only approximately 23% of GLOBE citizen science investigations had a fully formed and complex hypothesis (predictive, falsifiable, measurable relationship between more than one variable that could drive a complete investigation), whereas 53% of the student investigation had an oversimplified hypothesis or a hypothesis that was not fully formed; for example, such a hypothesis might predict the quantity/measure of a variable or simply propose measurements to explore a variable's range of values in the local environment. Finally, about 22% of project hypotheses contained a fundamental weakness or error, such as hypotheses that misunderstood the meaning of a key

variable, that based a prediction on a misunderstanding of the science content area, or that misunderstood the scope/scale of the hypothesis relative to the scope/scale of the investigation. This suggests that some additional attention may be needed to achieve citizen science goals related to understanding of the scientific process.

Most projects did not include a statistical analysis or included a poor/improper analysis (68%); in these projects, students might read values from a graph or table and visually interpret findings, while other students improperly fit lines or analyzed a different set of variables than those under investigation. It was quite common for students to plot the independent variable against time in one visualization or table, and to plot the dependent variable across time in a second visualization or table, rather than comparing them directly against each other, in cases where the research question was not about change over time. An additional 17% of projects included basic statistical analysis such as descriptive statistics, and 9% used more complex techniques such as inferential statistics and line fits. Similarly, in approximately 49% of projects, students created

a basic data representation such as a bar chart. Only 20% of projects demonstrated deeper understanding through data representations, such as projects that interpreted or described a pattern visible in the representation, or compared data values to a benchmark value.

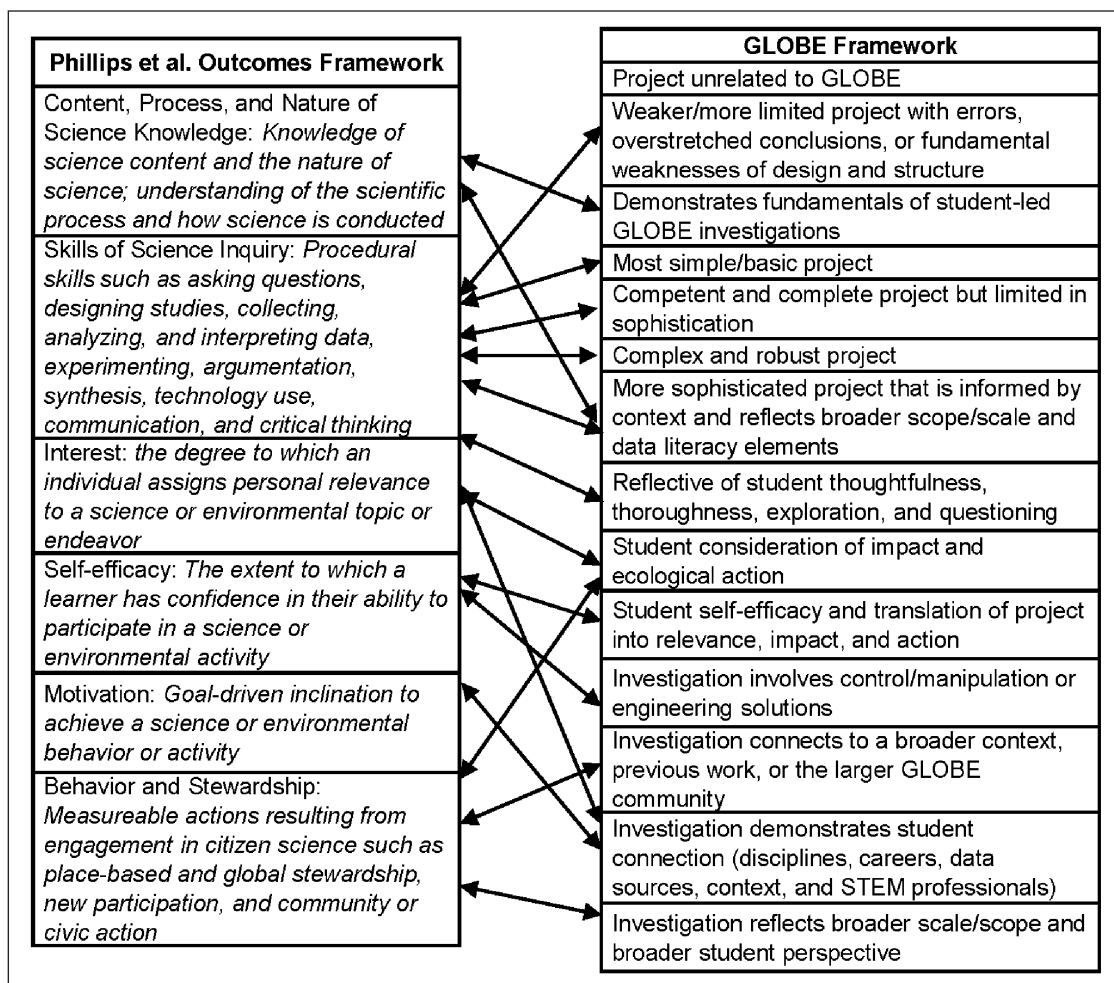
Although only 10% of GLOBE citizen science investigations were thoughtful about placing their smaller-scale project into the context of a major challenge or problem (such as climate change), most students did use their projects to respond to an impact on themselves, their school, or their community (58%). About one-quarter of projects proposed some actionable use of the knowledge or information they gained, while 14% did take action in the field, such as a community presentation or field site cleanup.

Taken together, these findings suggest that the GLOBE program has successfully encouraged citizen science projects to consider a broader perspective and has helped youth citizen scientists consider themselves as part of the environment that they study and the community to which they belong. Further action may be needed to translate

this deeper understanding to more robust investigation, analysis, and interpretation. The descriptive framework of clusters, discussed below, provides insight into areas where further work is needed.

### DISCUSSION OF THE CLUSTERS AND DESCRIPTIVE FRAMEWORK

Phillips et al. (2018) proposed a framework for measuring learning outcomes that does not exactly align to the purpose or the findings of this study, but does reflect the various ways that this study considered engagement and investigation characteristics. Rather than assessing individual outcomes, this GLOBE study considered how citizen scientists' developmental path along Phillips et al.'s outcomes like self-efficacy and skills of science inquiry would be reflected in the investigations they developed, conducted, and presented. **Figure 3** demonstrates some of the alignment between the two frameworks, and the ways in which multiple outcome types could be reflected in various elements of the GLOBE framework.



**Figure 3** Alignment between the GLOBE citizen science investigation framework and Phillips et al. (2018) framework for individual outcomes resulting from citizen science participation.

Similarly, it was interesting to consider co-occurrence of clusters within individual GLOBE student investigations. For instance, in 20 of the 36 projects (56%) that demonstrated manipulation of variables or engineering solutions, student citizen scientists also considered impact and ecological action, suggesting linkages between self-efficacy and action-oriented approaches to citizen science. Of the 15 investigations that indicated a sense of student self-efficacy and relevance of the project to their life, 10 (67%) conducted very simple, basic projects, suggesting that youth's skills in science inquiry may differ from their understanding of the value and relevance of the nature of science. Further investigation into the skills, identity, interests, motivations, and self-efficacy of youth participants in citizen science projects like GLOBE could further build connections between the outcomes of individual participants and the characteristics of the contributions they can make to citizen science initiatives.

We also observed patterns in how frequently student projects fell into each sub-cluster based on student age or grade band. The most simple/basic project sub-cluster was most common among lower primary students (40%). Similarly, the most complex/robust projects with more sophisticated indicators of science process skills were observed only in students at the middle school (10%) and secondary (18%) levels. By contrast, lower primary students were much more likely to consider impact and ecological action (60%, compared with 31% of projects overall) or to demonstrate change in their own self-efficacy (29%, compared with 7% overall). Lower primary students were also more likely to conduct investigations focused on engineering solutions or investigations with control variables (40% versus 17% overall).

## CONCLUSIONS AND FURTHER STUDY

These findings, in addition to serving in the GLOBE context for understanding the characteristics of student-driven citizen science investigations, provide a framework for typifying the experiences of youth as citizen scientists. A taxonomic tool based on this framework listing indicators for placing projects within the descriptive framework could provide understanding of a spectrum of projects. We propose that such a rubric be used to further investigate GLOBE and other citizen science investigations, particularly those that more heavily depend on citizen scientists driving their own experience, as opposed to scientist-driven contributory projects.

Our findings suggest multiple areas where additional support or intervention from the leaders of citizen science initiatives could benefit participants:

- 1. Unrelated projects:** In citizen science contexts in which community members are empowered to design and facilitate their own investigation and action projects, the results may be somewhat afield of initial intentions. GLOBE engages school-aged youth, and we observed a tendency among participants to present background research, like a book report. In the case of GLOBE, the program will likely continue to welcome any student contribution, but additional resources could help inspire student investigators to connect their interests to GLOBE's opportunities.
- 2. Questions lacking complexity:** Approximately half of the projects submitted to IVSS or SRS were centered on an overly simplified research question or hypothesis, such as projects that beg the question or ask isolated questions absent of context or content knowledge. Because of GLOBE's formal education context, it is unlike citizen science projects aimed at adults in that different levels of sophistication with scientific thinking, processes, and practices are developmentally appropriate. Many types of citizen science projects, however, hope to help their participants develop from their own starting point to more sophisticated conceptualizations of scientific inquiry. The taxonomy described in this paper may provide a starting point for scaffolding this development.
- 3. Needs around analysis and interpretation:** Although it is not surprising that citizen science investigations conducted by K–12 student participants would display a variety of difficulties in analyzing and interpreting data, these findings suggest specific focus areas. Many of the GLOBE investigations included no interpretation, or mis-used analytical tools; and 20% of the GLOBE projects worked directly with raw data collected by students in person, without translation into representations or models. Citizen scientists could benefit from checklists, templates, or guidance to help them select the best ways to display, manipulate, represent, and analyze their data depending on the type of question they have asked. In GLOBE's formal education context, this may be linked to the confidence and knowledge of teachers, who serve as guides through the investigation.

Citizen science initiatives could use the framework presented in this study as a formative assessment tool for understanding the experiences of citizen investigators, and identify areas of focus to support the learning and development of citizen science volunteers. Training and educational resources, experiences, and opportunities in key areas such as data analysis and interpretation could help move individual volunteers or group-led investigations toward higher levels of development. For GLOBE, this



study might lead to more robust future measurements of participant engagement and outcomes.

## DATA ACCESSIBILITY STATEMENT

To protect student investigator identities and privacy, raw data have not been made available. Supplemental files contain detailed information about processed, analyzed data. Readers should contact the corresponding author with any questions or requests.

## SUPPLEMENTAL FILES

The supplementary files for this article can be found as follows:

- **Supplemental File 1:** Frequency of Individual Code Occurrences by Project, Grade Band, and Country. xlsx (spreadsheet format). DOI: <https://doi.org/10.5334/cstp.480.s1>
- **Supplemental File 2:** Tabulation of Interrater Reliability Scores and Prevalence Indices.docx (document format). DOI: <https://doi.org/10.5334/cstp.480.s2>
- **Supplemental File 3:** Frequency of Project Alignment with Identified Clusters and Subclusters.xlsx (spreadsheet format). DOI: <https://doi.org/10.5334/cstp.480.s3>

## ETHICS AND CONSENT

As detailed in the text, the study was reviewed by the Oak Ridge Site-wide IRB (FWA 00005031) and was determined not to constitute human subjects research.

## ACKNOWLEDGEMENT

The authors would also like to thank the following individuals for their significant assistance: Jen Bourgeault and Haley Wicklein (University of New Hampshire) and Janet Struble (University of Toledo) provided photography and other data for SRS. Other members of the GLOBE Implementation Office, including Amy Barfield and Sarah Parsons, provided insights on the meaning of the findings for GLOBE.

## FUNDING INFORMATION

This work was supported by NASA through a cooperative agreement to Oak Ridge Associated Universities (Grant # 80NSSC18K1589) and an award to UCAR/NCAR for the

GLOBE Implementation Office (Grant # 80NSSC19M0120), and by both NASA and NSF through funding for the SRS (NASA: #80NSSC18K0135; NSF: 1546713). ORAU supported the interrater reliability analysis through a 2019 Thought Leadership Research Award.

## COMPETING INTERESTS

LC was the NASA GLOBE Program Manager and initiated the funding that supported this study; she currently manages the NASA Science Mission Directorate Science Activation program which supports the GLOBE Observer app. KC directs the University of Toledo's GLOBE partnership and is funded by NASA to support teachers and students involved in GLOBE; he has mentored students who presented investigations in the GLOBE IVSS. JM worked for The GLOBE Program, leading the development and management of the IVSS, and received funding from NASA during the writing of this manuscript. AM was funded by NASA through a cooperative agreement (80NSSC18K1589) to complete this GLOBE-related study while employed by Oak Ridge Associated Universities, and has previous involvement in conducting evaluations of GLOBE initiatives. KM-B has no competing interests to disclose.

## AUTHOR CONTRIBUTIONS

Author Martin developed the study and research questions, ensured protection of human subjects, and analyzed/coded data and prepared findings. Authors Chambers, Czajkowski, and Malmberg provided insights, reviews, and code development for the list of codes, and Malmberg contributed substantially to the preparation of this manuscript. Author Miller-Bains served as the additional/independent coder in the interrater reliability analysis, completed that analysis, and contributed to this manuscript. All authors reviewed codes/findings, contributed to and reviewed this manuscript before submission.

## AUTHOR AFFILIATIONS

**Ann Martin**  [orcid.org/0000-0003-3201-2041](https://orcid.org/0000-0003-3201-2041)  
Oak Ridge Associated Universities, US

**Katherine Miller-Bains**  [orcid.org/0000-0002-0752-1461](https://orcid.org/0000-0002-0752-1461)  
University of Virginia, US

**Julie Malmberg**  
University Corporation for Atmospheric Research, US

**Lin Chambers**  [orcid.org/0000-0002-9509-7022](https://orcid.org/0000-0002-9509-7022)  
National Aeronautics and Space Administration, US

**Kevin Czajkowski**  [orcid.org/0000-0002-0472-4204](https://orcid.org/0000-0002-0472-4204)  
University of Toledo, US



## REFERENCES

- Aristeidou, M** and **Herodotou, C.** 2020. Online citizen science: A systematic review of effects on learning and scientific literacy. *Citizen Science: Theory and Practice*, 5(1): 1–12. DOI: <https://doi.org/10.5334/cstp.224>
- Bastian, M, Heymann, S** and **Jacomy, M.** 2009. Gephi: an open source software for exploring and manipulating networks. In: *Third International AAAI Conference on Weblogs and Social Media*. San Jose, California on 17–20 May 2009.
- Blondel, VD, Guillaume, J-L, Lanbiotte, R** and **Lefebvre, E.** 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10: 10008. DOI: <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bonney, R, Ballard, H, Jordan, R, McCallie, E, Phillips, T, Shirk, J** and **Wilderman, CC.** 2009. *Public participation in scientific research: Defining the field and assessing its potential for informal science education. A CAISE Inquiry Group report*. Washington, DC: Center for Advancement of Informal Science Education (CAISE).
- Chase, SK** and **Levine, A.** 2016. A framework for evaluating and designing citizen science programs for natural resources monitoring. *Conservation Biology*, 30(3): 456–466. DOI: <https://doi.org/10.1111/cobi.12697>
- Craven, J** and **Hogan, T.** 2008. Rethinking the science fair. *Phi Delta Kappan*, 89(9): 679. DOI: <https://doi.org/10.1177/003172170808900915>
- DeLisi, J, Kook, JF, Levy, AJ, Fields, E** and **Winfield, L.** 2020. An examination of the features of science fairs that support students' understandings of science and engineering practices. *J Res Sci Teach*, 58(4): 491–519. DOI: <https://doi.org/10.1002/tea.21669>
- Doerfel, ML.** 1998. What constitutes semantic network analysis? A comparison of research and methodologies. *Connections*, 21(2): 16–26.
- Eugenio, BD** and **Glass, M.** 2004. The kappa statistic: A second look. *Computational Linguistics*, 30(1): 95–101. DOI: <https://doi.org/10.1162/089120104773633402>
- Finarelli, MG.** 1998. GLOBE: A worldwide environmental science and education partnership. *Journal of Science Education and Technology*, 7: 77–84. DOI: <https://doi.org/10.1023/A:1022588216843>
- Frensley, T, Crall, A, Stern, M, Jordan, R, Gray, S, Prysby, M, Newman, G, Hmelo-Silver, C, Mellor, D** and **Huang, J.** 2017. Bridging the benefits of online and community supported citizen science: A case study on motivation and retention with conservation-oriented volunteers. *Citizen Science: Theory and Practice*, 2(1): 4, 1–14. DOI: <https://doi.org/10.5334/cstp.84>
- Geoghegan, H, Dyke, A, Pateman, R, West, S** and **Everett, G.** 2016. *Understanding motivations for citizen science*. Reading, UK: UKEOF, University of Reading, Stockholm Environment Institute (University of York) and University of the West of England.
- Houseal, AK, Abd-El-Khalick, F** and **Destefano, L.** 2014. Impact of a student–teacher–scientist partnership on students' and teachers' content knowledge, attitudes toward science, and pedagogical practices. *Journal of Research in Science Teaching*, 51(1): 84–115. DOI: <https://doi.org/10.1002/tea.21126>
- Kastens, K.** 2014a. *Pervasive and persistent understandings about data*. Boston, MA: EDC Oceans of Data Institute. Available at <http://oceansofdata.org/sites/oceansofdata.org/files/pervasive-and-persistent-understandings-01-14.pdf>.
- Kastens, K.** 2014b. *The relationship between direct and data-mediated knowledge of the world*. Boston, MA: EDC Oceans of Data Institute. Available at <http://oceansofdata.org/sites/oceansofdata.org/files/representation-and-referent-01-14.pdf>.
- Malmberg, J** and **Maul, K.** 2013. Supporting climate science research with 21<sup>st</sup> century technologies and a virtual student conference for upper elementary to high school students. *LEARNing Landscapes*, 6(2): 249–264. DOI: <https://doi.org/10.36510/learnland.v6i2.615>
- Minner, DD, Levy, AJ** and **Century, J.** 2010. Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4): 474–496. DOI: <https://doi.org/10.1002/tea.20347>
- National Academies of Sciences, Engineering and Medicine.** 2018. *Learning through citizen science: Enhancing opportunities by design*. Washington, DC: The National Academies Press. DOI: <https://doi.org/10.17226/25183>
- National Advisory Council for Environmental Policy and Technology.** 2016. *Environmental protection belongs to the public: A vision for citizen science at EPA*. Washington, DC: Environmental Protection Agency Report 219-R-16-001, National Service Center for Environmental Publications. Available at <https://www.epa.gov/faca/nacept-2016-report-environmental-protection-belongs-public-vision-citizen-science-epa>.
- Phillips, T, Porticella, N, Conostas, M** and **Bonney, R.** 2018. A framework for articulating and measuring individual learning outcomes from participation in citizen science. *Citizen Science: Theory and Practice*, 3(2): 3. DOI: <https://doi.org/10.5334/cstp.126>
- Phillips, TB, Ballard, HL, Lewenstein, BV** and **Bonney, R.** 2019. Engagement in science through citizen science: Moving beyond data collection. *Science Education*, 103(3): 665–690. DOI: <https://doi.org/10.1002/sce.21501>
- Prell, C.** 2012. *Social Network Analysis: History, Theory and Methodology*. Thousand Oaks, CA: SAGE.
- R Core Team.** 2021. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Saldaña, J.** 2015. *The Coding Manual for Qualitative Researchers*. 2nd ed. Thousand Oaks, CA: SAGE.

**The GLOBE Program.** 2018a. *The GLOBE Program strategic plan 2018–2023*. Boulder, CO: University Corporation for Atmospheric Research. Available at <https://www.globe.gov/documents/10157/32400137/GLOBE+Strategic+Plan+2018.pdf/e4a03152-45b6-4554-95a7-a0dec961d806>.

**The GLOBE Program.** 2018b. *International Virtual Science Symposium – rubrics*. Boulder, CO: University Corporation for Atmospheric Research. Available at <https://www.globe.gov/news-events/globe-events/virtual-conferences/2018-international-virtual-science-symposium/rubrics>.

---

**TO CITE THIS ARTICLE:**

Martin, A, Miller-Bains, K, Malmberg, J, Chambers, L and Czajkowski, K. 2022. Characterizing Student-Driven Research Investigations Contributed to the GLOBE Program Citizen Science Initiative in a Formal Education Context. *Citizen Science: Theory and Practice*, 7(1): 11, pp. 1–16. DOI: <https://doi.org/10.5334/cstp.480>

Submitted: 05 December 2021 Accepted: 22 February 2022 Published: 24 March 2022

**COPYRIGHT:**

© 2022 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 <http://creativecommons.org/licenses/by/4.0/>.

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

