



# Addressing Common Student Technical Errors in Field Data Collection: An Analysis of a Citizen-Science Monitoring Project

Joanna Philippoff<sup>1\*</sup> and Erin Baumgartner<sup>2</sup>

<sup>1</sup>*Curriculum Research & Development Group, College of Education, University of Hawai'i at Mānoa, Honolulu, HI 96822,*

<sup>2</sup>*Department of Biology, Western Oregon University, Monmouth, OR 97361*

**The scientific value of citizen-science programs is limited when the data gathered are inconsistent, erroneous, or otherwise unusable. Long-term monitoring studies, such as Our Project In Hawaii's Intertidal (OPIHI), have clear and consistent procedures and are thus a good model for evaluating the quality of participant data. The purpose of this study was to examine the kinds of errors made by student researchers during OPIHI data collection and factors that increase or decrease the likelihood of these errors. Twenty-four different types of errors were grouped into four broad error categories: missing data, sloppiness, methodological errors, and misidentification errors. "Sloppiness" was the most prevalent error type. Error rates decreased with field trip experience and student age. We suggest strategies to reduce data collection errors applicable to many types of citizen-science projects including emphasizing neat data collection, explicitly addressing and discussing the problems of falsifying data, emphasizing the importance of using standard scientific vocabulary, and giving participants multiple opportunities to practice to build their data collection techniques and skills.**

## INTRODUCTION

Citizen science involves engaging cadres of volunteers in scientific research. An increasingly popular research approach, citizen science enables the collection of large-scale datasets across time and space and can play a valuable role in conducting ecological research because it allows the scientific community to address questions that it would otherwise be unable to for logistical or financial reasons (12). Citizen scientists are not professional scientists; they may be volunteers connected to a particular wildlife area or visitors to a website who become intrigued by the opportunity to help collect or examine data. Citizen-science projects provide opportunities for the public to engage in the process of science through authentic scientific experiences, contributing to a more scientifically literate society (10).

Citizen science can be a valuable instructional model in K–12 and postsecondary classrooms. However, the educational opportunities afforded by citizen science projects need to be balanced with the importance of collecting scientifically useful data (8). The adoption of citizen-scientist data

to examine scientific issues can be hindered by the perception that the data are not reliable (7). When questions arise, citizen-science data are underutilized, which is detrimental to both scientific and educational goals. However, there is mounting evidence supporting the use of citizen-science data from a variety of projects. For example, comparisons between datasets collected by volunteer citizen scientists and those of professional scientists have shown consistent rates of error and bias (3). Volunteers with training, even limited training, are able to collect reliable data when provided with unambiguous standardized protocols (9).

Our Project In Hawaii's Intertidal (OPIHI) is a citizen science program in which middle and high school students survey rocky intertidal areas in Hawai'i (1, 2). To address reservations scientists may have had about using OPIHI student-generated data, we conducted a validity assessment that demonstrated students' data quality was robust and similar to that of more experienced professional researchers (5). OPIHI student data resulted in the first description of community-level patterns at multiple intertidal sites across the Hawaiian islands (6).

The purpose of this paper is not to determine the ability of students and other citizen scientists to conduct scientific research, which OPIHI and many other projects have already demonstrated, but rather to determine the types of data collection errors made by students involved in a monitoring project. We categorized and analyzed the frequency of

\*Corresponding author. Mailing address: Curriculum Research & Development Group, College of Education, University of Hawai'i at Mānoa, 1776 University Ave., Honolulu HI, 96822. Phone: 808-956-4951. E-mail: philippo@hawaii.edu.

occurrence of different types of errors committed during data collection in order to enhance instruction, minimize error, and improve data quality. This focus on prior training does not remove the need for post hoc data filtration procedures in citizen science, such as expert review, but is an important aspect of a robust project improvement cycle. The objectives of this study were to 1) determine the types of data collection errors made by student researchers, 2) determine what factors increase or decrease the likelihood of these errors, and 3) suggest ways to strategize instruction and refine protocols to reduce citizen science data collection errors in the field.

## PROCEDURE

We examined the quality of student data sheets from 47 OPIHI field trips. Trips were conducted by nine teachers at eight different schools to 13 different intertidal sites from February 2004 to May 2007. Participating students, from grades 5 to 12, worked in groups with a volunteer chaperone while collecting data. Prior to the field trips, teachers led students in core OPIHI curricula, including lessons on field methodology and species identification (for a full description of OPIHI methods, activities, and protocols see 1 and 5).

OPIHI protocols use traditional ecological sampling methods including transects and quadrats. Students completed separate data sheets on waterproof paper for each of up to three different sampling methods along a transect line. We analyzed student data sheets from 233 transect lines. Data sheets included the scientific names of the most abundant intertidal organisms as well as blank spaces for students to write in the names of additional organisms. Because students worked in groups, and different teachers focused on different sampling methods, to avoid pseudoreplication of data one unit of analysis encompassed all of the sampling data collected along a transect. Although each specific error type could occur multiple times on each data sheet, transect groups were considered to have either committed the error (“present”) or not (“absent”). Data collection procedures in citizen-science projects that emphasize recording the presence of organisms, rather than the absence, have been shown to be more reliable than projects that rely on opportunistic data collection (13).

OPIHI data sheets followed a standard format, although teachers often modified them to streamline data collection when trips had limited time, when students required differentiation, or to highlight particular classroom content or elements of instruction. Because of these modifications, not all errors were possible on each data sheet. For example, some teachers filled in location information prior to photocopying, thus the error of location omission would not be possible on these data sheets. To correct for different potential numbers of mistakes students on each field trip could make, we calculated the percent error of each specific type of data collection error.

The relationship between student error rate and teacher experience, assistant experience, student age and experience, and instructional time was examined using Pearson correlation coefficients and independent-samples *t*-tests. Experience was defined as the number of OPIHI field trips in which each citizen scientist participated. Information on instructional time was collected in 2007 by surveying teachers to determine how much time they had spent in class on OPIHI curricula.

## RESULTS

We identified 24 different types of specific errors that were grouped into four broad categories (Table 1). Of these broad categories, “sloppiness” was the most prevalent error type. “General category,” a specific error in the sloppiness group, had the highest single error type—over 40% of student groups had at least one instance of this error on their data sheets. An example of a “general category” error would be writing “urchin” on the data sheet as opposed to the specific species of urchin, e.g., “*Echinometra mathaei*.” Using the scientific names for organisms, to a taxonomic level appropriate for field identification and student age, is important in biodiversity monitoring studies to accurately assess trends in ecological patterns.

Over one-third of student groups had quadrat totals that did not add up to the number of intercepts utilized in OPIHI protocols, or 100% cover. Alarming, we found these students recorded that their data accurately totaled the correct coverage, indicating they never added their intercepts and falsified the tally. We hypothesize students may have felt too rushed in the field to complete this arithmetic “data check”; further evidence to support this hypothesis is the high prevalence of messy writing errors. Another error with a high rate of occurrence in the sloppiness category was the incorrect or suspicious substrate error (e.g., recording “limestone” for a site known to be a basalt bench). This error, along with high rates of incorrectly recoding multiple species under a single transect point (a methodology error), may indicate inadequate preparation for reporting certain site characteristics and the need for additional instruction and practice using OPIHI methods.

Error rates decreased with OPIHI teacher experience ( $r = -0.20$ ,  $n = 232$ ,  $p < 0.01$ ) and volunteer assistant experience; volunteers who had previously facilitated a field trip had significantly fewer errors on their mentee data sheets,  $t(42) = 3.35$ ,  $p < 0.01$ . Older students made fewer errors than younger students ( $r = -0.34$ ,  $n = 232$ ,  $p < 0.01$ ). Interestingly, there was no discernable relationship between OPIHI class instructional time before field trips (range 6.5 to 29 hours) and data sheet error rate. However, instructional time spent on sampling methods (5 to 10 hours) was more consistent than time spent on learning intertidal identifications (0 to 22 hours) prior to OPIHI field trips. Although not significant, overall error rates decreased with the number of field trips students went on. However, the percentage of

TABLE 1.  
Specific student data sheet errors grouped into four broad categories.

Error Category	Specific Errors <sup>a</sup>	N Transects	Percent Error (%)
Sloppiness	General category utilized (e.g., “urchin”)	233	41.6
	Quadrat cover did not add to the correct number of intercepts or percent (100%) and final tally falsified	211	35.6
	Incorrect substrate and/or incorrect or suspicious substrate size	102	34.3
	Writing so messy not discernable	233	31.3
	Quadrat “ticks” did not add to the correct number of intercepts and never tallied	211	31.3
	Genus but no species name (or species name with no genus)	233	24.5
	Suspicious wind speed	129	14.7
	Quadrat “ticks” never tallied (but do add to correct number of intercepts)	211	8.1
	Missing Data	No start and/or end times	39
No (or incomplete) site conditions		182	27.5
No transect number (metadata geographical information)		226	20.4
No location		110	13.6
No date		200	9.5
No names		213	8.0
Methodological	Multiple records at transect intercepts	208	30.8
	Wrong method	233	7.7
	Transient objects recorded as data point (e.g., leaf litter or trash)	233	6.4
	“Water” recorded as data point	233	2.2
Misidentification	Misidentification or suspicious species identification	233	18.5
	Suspicious amount of space given to small organisms	233	3.9

<sup>a</sup> Specific errors are in order of frequency, with the most frequent errors in each category listed first.

missing data errors increased. This may be due to students not understanding the importance of research metadata, such as recording start and end times of sampling.

### Implications for Practice

Based on our systematic analysis of data sheets from a successful monitoring project, we suggest a number of strategies to reduce citizen-science data collection errors. Some of our recommendations will be most relevant to students participating in course-based research, including those at the postsecondary level, as they have implications for implementing active-learning strategies (4). Other recommendations are applicable to any citizen-science program that features a training component. For example, several of the most common errors we identified, particularly those regarding sloppy or falsified data, are likely unique to situations where participation is required as a course component—in contrast to citizen-science projects whose participants are self-selected interested volunteers.

Student motivation and enthusiasm for accuracy may be reduced and data collection rushed due to course structure. Other types of errors, such as identification errors, are found across a number of citizen-science projects; our suggestions for decreasing these errors align with those put forth by other researchers (8).

**Emphasize neat data collection.** In school settings, students can examine the readability of data sheets from previous years or swap data sheets when inputting information; neatness can be included when assessing student performance. Encourage citizen scientists to identify potential sources of error and recommend corrections—implement their suggestions to continually refine and improve data collection procedures. In addition to reducing errors, participant engagement may be increased when they have the opportunity to help guide data collection and entry protocols. Consider the use of technology to help minimize sloppy errors if appropriate to your project and environment.

**Explicitly address and discuss the problem of falsifying data.** Examining the validity and reliability of volunteer data is good practice for all citizen-science projects. However, these practices often occur after sampling. To facilitate conscientious data collection, discuss the potential scientific and ethical implications of cutting corners as part of project training. Cultivating a classroom or community culture that is collaborative, as opposed to competitive, may decrease these types of errors and allow for instructive discussions should concerns arise.

**Practice data collection techniques.** In this study, practice appeared to be more effective than direct instruction at building data collection skills. Ecological data collection can be practiced in the field—or even in a classroom or courtyard before a field trip. Other citizen science projects have found error rates decrease with modest training (9); reviews of successful projects show practice is particularly effective for studies requiring taxonomic identification (8). For K–12 and postsecondary students, practice is also a component of enhanced learning. Continued opportunity to engage in data collection protocols not only improves scientific accuracy, it builds scientific content and skills knowledge (11).

**Practice identification skills.** We hypothesize one of the reasons for the large number of “general category” errors in this study is that both teachers and students are uncomfortable using scientific names. Emphasizing the importance of using common language, in particular scientific terms appropriate to the citizen-science project, helps to standardize data collection and thus can enhance scientific usage.

**Tailor data sheets and identification cards to sites.** Opportunistic data collection by volunteers can be particularly subject to bias (13). In this study, we found students prioritized species identifications based on the most easily accessible information. To address these errors, students or volunteers can be recruited to refine data sheets or build and refine field guides so information is easily accessible and user-friendly, but the amount is not overwhelming, as it is site-specific.

**Recruit experienced volunteers to assist with the data collection process.** Depending on your project, volunteers can be adult assistants, scientists, or more experienced peers. Students who have already completed a course can serve as mentors. At the postsecondary level, independent or work study opportunities are ways to recruit a cadre of more experienced volunteers.

These approaches to reducing errors in data collection emphasize the importance of the role of an instructor or lead facilitator in enhancing citizen-science data quality. Error rates may be indicative of how careful students or volunteers were during data collection and thus reflective

of overall data quality. Being aware of common data collection errors allows educators and trainers to anticipate and address potential pitfalls in data collection.

## ACKNOWLEDGMENTS

This research was supported in part by a National Science Foundation Graduate Fellowships in K–12 Education grant (DGE02-32016) and was approved by the University of Hawai‘i Committee on Human Subjects (CHS #15657). We would like to thank T. Erin Cox, Chela Zabin, and Kanesa Duncan Seraphin for intellectual and programmatic contributions, the teachers and students who collected data, and the graduate students who volunteered as research assistants on field trips. The authors declare that there are no conflicts of interest.

## REFERENCES

1. **Baumgartner, E., and C. Zabin.** 2008. A case study of project-based instruction in the ninth grade: a semester-long study of intertidal biodiversity. *Environ. Educ. Res.* **14**:97–114.
2. **Baumgartner, E., C. Zabin, J. Philippoff, T. E. Cox, and M. Knope.** 2009. Ecological monitoring provides a thematic foundation for student inquiry, p 191–209. *In* Yagar, R., (ed), *Inquiry: The Key to Exemplary Science*. National Science Teachers Association Press, Arlington, VA.
3. **Bird, T. J., et al.** 2014. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* **173**:144–154.
4. **Brownell, S. E., and M. J. Kloser.** 2015. Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology. *Stud. Higher Educ.* **40**(3):525–544.
5. **Cox, T. E., J. Philippoff, E. Baumgartner, and C. M. Smith.** 2012. Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecol. Appl.* **22**(4):1201–1212.
6. **Cox, T. E., J. Philippoff, E. Baumgartner, C. Zabin, and C. M. Smith.** 2013. Spatial and temporal patterns of intertidal communities along the main Hawaiian islands. *Pacific Sci.* **67**(1):23–45.
7. **Foster-Smith, J., and S. M. Evans.** 2003. The value of marine ecological data collected by volunteers. *Biol. Conserv.* **113**:199–213.
8. **Freitag, A., and M. J. Pfeffer.** 2013. Process, not product: investigating recommendations for improving citizen science “success.” *PLOS one.* **8**(5):1–5.
9. **Fuccillo, K., T. Crimmins, C. de Rivera, and T. Elder.** 2015. Assessing accuracy in citizen science-based plant phenology monitoring. **59**(7):917–926.
10. **Lovell, S., M. Hamer, R. Slotow, and D. Herbert.** 2009. An assessment of the use of volunteers for terrestrial invertebrate biodiversity surveys. *Biodivers. Conserv.* **18**:3295–3307.
11. **Orr, T., and E. Baumgartner.** 2012. Participatory science improves scientific literacy in pre-service elementary teachers.

Proceedings of the National Association of Biology Teachers Four-Year College and University Section 2012 Research Symposium. [Online.] [http://www.nabt.org/websites/institution/File/docs/Four%20Year%20Section/2012%20Proceedings/Orr %20and%20Baumgartner%20NABT%20symposium.pdf](http://www.nabt.org/websites/institution/File/docs/Four%20Year%20Section/2012%20Proceedings/Orr%20and%20Baumgartner%20NABT%20symposium.pdf).

12. **Silvertown, J.** 2009. A new dawn for citizen science. *Trends Ecol. Evol.* **24**:467–472.
13. **van Strien, A. J., C. A. M. van Swaay, and T. Termaat.** 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analyzed with occupancy models. *J. Appl. Ecol.* **50**:1450–1458.