Cross-Project Analysis of Volunteers' Scientific Observation Skills

VERONICA DEL BIANCO (iD)

KAREN PETERMAN (iD)

RACHEL BECKER-KLEIN (iD)

CATHLYN M. DAVIS (iD)

ANDREA GROVER (iD)

TINA PHILLIPS (iD)

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

ABSTRACT

This paper explores the assumptions that citizen science (CS) project leaders had about their volunteers' science inquiry skill–proficiency overall, and then examines volunteers' actual proficiency in one specific skill, scientific observation, because it is fundamental to and shared by many projects. This work shares findings from interviews with 10 project leaders related to two common assumptions leaders have about their volunteers' skill proficiency: one, that volunteers can perform the necessary skills to participate at the start of a CS project, and therefore may not need training; and two, volunteer skill proficiency improves over time through involvement in the CS project. In order to answer questions about the degree of accuracy to which volunteers can perform the necessary skills and about differences in their skill proficiency based on experience and data collection procedures, we analyzed data from seven CS projects that used two shared embedded assessment tools, each focused on skills within the context of scientific observation in natural settings: *Notice relevant features for taxonomic identification* and *record standard observations*. This across-project and cross-sectional study found that the majority of citizen science volunteers (n = 176) had the necessary skill proficiency to collect accurate scientific observations but proficiency varied based on volunteer experience and project data collection procedures.

INTRODUCTION

Citizen science (CS) engages the public in scientific research, and thus it is critical that volunteers on CS projects be proficient in project-specific science inquiry skills in order for them to contribute high-quality data and to meet both learning and scientific outcomes (National Academies of Sciences, Engineering, and Medicine 2018). The National Research Council (NRC) has defined science inquiry skills, broadly, as all tasks required to pursue the work of science (NRC 2012), and these skills are a defining attribute used to characterize models of CS (Bonney et al. 2009; Shirk et al. 2012; Phillips et al. 2019). Assessment of volunteers' proficiency with science inquiry skills can lead to improvements in data quality and confidence in volunteers' efforts while supporting CS projects' learning goals (Davis et al. 2022).

Despite this, recent research has documented a clear misalignment between what the field of CS says about the fundamental importance of skilled volunteers and assessment efforts to ensure volunteers have those necessary skills (Bowser et al. 2020; Burgess et al. 2017; Phillips et al. 2018; Stylinski et al. 2020). For example, in their review of 327 CS projects' stated goals and objectives, Phillips et al. (2018) found the majority (59%) centered on influencing volunteers' skills related to data collection and monitoring, implying the importance of said skills. Yet, of the 72 projects that responded to the questions about evaluation, only 28% measured outcomes related to skills—making it the lest measured outcome across all projects. Similarly, only 4 of the 36 CS project proponents (either scientific leads or data managers) interviewed by Bowser at al. (2020) conducted volunteer testing or skill assessment, and Burgess et al. (2017) found of the 125 CS projects they surveyed, 30.9% conducted post-tests, which provide a measure of confidence in data collectors' abilities. Possible barriers to assessing these skills include lack of time, of staff, of expertise, of funding, and of supporting resources (Stylinski et al. 2020).

Given limited assessment efforts, it is not surprising that published research on volunteers' proficiency with science inquiry skills is limited in scope and depth. No articles about what project leaders think about their volunteers' skill proficiency were found. The majority of skill assessments around collection procedures are self-reported by participants and are project-specific (Philips et al. 2018; Peter et al. 2019; National Academies of Sciences, Engineering, and Medicine 2018; Stylinski et al. 2020; Peterson et al. 2022). These studies have found that most volunteers can perform basic skills, such as observing species, identifying species, and collecting data in a standardized manner (Crall et al. 2011; Haywood et al. 2016; Peter et al. 2021).

In a systematic literature review of participant outcomes of biodiversity CS projects, Peter et al. (2019) found six studies that addressed the acquisition of new scientific skills.

Among these, only two studies investigated skill gains across multiple biodiversity CS projects; however, the two papers did not specify the kind of scientific skill gains, but instead only reported they were positive changes (Bela et al. 2016; Fernandez-Gimenez, Ballard, and Sturtevant 2008). The lack of specificity and reliance on self-reported outcomes can bias toward socially desirable answers (Furnham 1986), and suggests such results should be interpreted conservatively.

Analyses of assessments across projects provide insight into broad trends associated with volunteers' skill proficiency, while direct measures of skills (i.e., beyond self-reported measures) are important complements to self-reported volunteer data and can help validate the quality of data collected by volunteers (Fu, Kannan, and Shavelson 2019). Thus, direct measures of skills across multiple projects would be particularly useful to expand both the scope and depth of research on volunteers' proficiency with science inquiry skills.

THE PRESENT RESEARCH

Funded by the National Science Foundation (DRL #1713424), our team studied volunteer skill assessment processes and impacts within the context of CS and published a series of articles based on the results. All of the papers center on the development and use of embedded assessment (EA) and activities integrated into the learning experience that allow learners to demonstrate their science competencies for assessment purposes (Becker-Klein et al. 2023; Davis et al. 2022; Peterman et al. 2022; Stylinski et al. 2020).

Within the context of this larger study, we worked collaboratively with staff of 10 CS projects to identify and articulate science inquiry skills common across the projects, and then to develop and implement assessment measures for those skills that could be used by more than one project (see Becker-Klein et al. 2023). Here, we examined findings from the projects' implementation of these shared measures to explore volunteers' scientific observation skills in natural settings. Our definition of this overarching skill is based in Eberbach and Crowley's (2009) paper on how scientific observation evolves. As they wrote, "Scientific observation is not a domain-general practice, but one that goes hand in hand with disciplinary knowledge, theory, and practice" (p. 41), and is in contrast to "everyday observations as those that occur with little or no knowledge of the constraints and practices of scientific disciplines" (p. 46).

BREAKING DOWN SCIENCE INQUIRY SKILLS FOR ASSESSMENT

Because science inquiry skills are so broad, it is necessary to move beyond simple lists (e.g., observation, exploration,

questioning, prediction, experimentation, argumentation, interpretation, and synthesis) and to conceptualize skills in practice and from the perspective of the volunteer. Likewise, attention and intention are needed to break down broad ideas about science inquiry skills into the smaller, tangible, and measurable underlying dimensions. In our research, we broke down scientific observation into two skills.

The first skill assessed is *notice relevant features*. Eberbach and Crowley (2009, p. 43) define "noticing" as "[using] existing knowledge to notice and organize key features that support inferences about deep principles and relationships within biological systems." That is, an observer is able to match what they see with their knowledge of disciplinary structure. Often, relevant features are used either to distinguish the animal from the background environment (e.g., in a photo) or to accurately identify the organism at a prescribed taxonomic level, such as species.

The second skill assessed is *record standard observations*. Eberbach and Crowley (2009, p. 56) define this skill as "record[ing] observations using established disciplinary procedures, standards, and representations." Standard observations can be spatial (e.g., GPS coordinates), temporal (e.g., date and time of day), environmental (e.g., cloud cover, ground cover), or biological (e.g., species identification or percent of tree leaf senescence).

Volunteers may vary in their proficiency at these skills based on volunteer experience and project data collection procedures. Furthermore, CS project leaders may harbor misconceptions about volunteer skill proficiency, such as assuming that the extent of participation results in improvements. To understand proficiency and associated factors, we asked the following questions:

1. What do project leaders think about their volunteers' skill proficiency?
2. To what extent are volunteers proficient at dimensions of scientific observation?
3. What differences exist in volunteer skill proficiency based on volunteer experience and project data collection procedures?

METHODS

For this research, data was collected from two different groups (CS project leaders and CS volunteers) using two distinct approaches (interviews and embedded assessments, respectively). The follow sections explain the methods used with each group of participants.

PROJECT LEADER PARTICIPANTS
Data collection
The 10 CS project leaders in this study participated in three semi-structured interviews via the Zoom video conferencing platform and two in-person meetings. They were selected for our study because all their projects focused on the skill of scientific observation, and are representative of the variety in CS projects, incorporating monitoring, species identification, and image classification. The purpose of the first interview in the fall of 2018 was to gather initial information about project activities and targeted science inquiry skills. The two in-person meetings, in December 2018 and in December 2019, focused on the development of the shared embedded assessments. All interviews lasted between 30 and 60 minutes and were recorded and later transcribed verbatim. Project leaders were compensated annually for three years with stipends for their participation in the overall NSF research project, which included the development and implementation of the embedded assessment into their CS projects.

Coding
To analyze the interview and meeting notes, three researchers used a six-step collaborative qualitative analysis (Richards and Hemphill 2018) to develop an inductive scheme documenting the assumptions that CS project leaders had about their volunteers' skill proficiency. Interview and in-person meeting notes were then coded and analyzed using NVivo12. Using consensus coding, two researchers coded each interview independently and then compared codes; all disagreements were discussed, and the final codebook was agreed upon (see Appendix A).

VOLUNTEER PARTICIPANTS
Data collection
The volunteers involved in this study were recruited by the CS project leaders who were implementing embedded assessments of the two skills. In total, 176 volunteers from seven CS projects participated in the embedded assessments between July 2019 and October 2020. Note, the three remaining projects did not implement these particular assessments, and so did not have any volunteer data to include in this part of the study. Volunteers were not compensated because the assessments were embedded into their projects and were not an additional burden to complete.

Seventy-eight volunteers from five citizen science projects participated in the *record standard observations* embedded assessment. All five projects used data collection procedures that ask volunteers to record at least three categories of standard observations (including

spatial, temporal, environmental, and biological) for organisms (plants and animals).

Ninety-eight volunteers from three CS projects participated in the *notice relevant features for taxonomic identification* embedded assessment. In these three projects, the data collection procedures ask volunteers to identify animals (insects and mammals) to the lowest possible taxonomic rank.

Embedded assessment instruments

The instruments used to assess these volunteer skills were co-developed with the participating CS project leaders. These embedded assessments are meant to determine whether an adult volunteer can accurately *notice relevant features for taxonomic identification* or *record standard observations,* not whether they do so consistently within the parameters of the CS project (see Becker-Klein et al. 2023 for more details on the development and validation of these assessments). Brief descriptions of each instrument are included below:

- *Notice relevant features.* This assessment presents volunteers with images that replicate typical photos of organisms taken in the field by volunteers or from camera traps. Assessment participants are then asked to identify the organism(s) they see in the photos just as they are in the actual protocol, and list the relevant features they noticed in their identification process.
- *Record standard observations.* This assessment presents a video clip to simulate the first-person perspective of a volunteer collecting field data. Assessment participants are then asked to record the standard observations (i.e., date, location, ground cover) that are requested on the project's data sheet.

Assessment scoring

The embedded assessments administered by the seven CS projects were originally scored by the respective CS project leaders (see Supplemental File 1: Appendices A–C for example scoresheets). One of our team members verified scores and then aggregated the data in Excel by measure: *notice relevant features* and *record standard observations.* Missing data were coded as zero. Volunteers participating in their first field season were coded as new volunteers, and volunteers who participated in two or more seasons were coded as returning volunteers.

Data analysis

To explore the extent of volunteers' ability to *notice relevant features for taxonomic identification,* and *record standard observations,* a percentage of accurate answers was calculated for each. Three independent t-tests were conducted to determine if there was (1) a difference in accuracy of species identification between volunteers who could notice correctly one or more relevant features of an organism and those that could not, (2) a difference in ability to accurately *notice relevant features* and identify species between new and returning volunteers, and (3) a difference in volunteer accuracy to *record standard observations* of animals versus plants. Cohen's d was calculated for each t-test to determine the effect size for the comparison between the two means.

RESULTS

As previously stated, data was collected from two different groups, CS project leaders and CS volunteers, and analyzed separately. The following sections explain the results from the data analysis from each group of participants.

PROJECT LEADER ASSUMPTIONS ABOUT SKILLS

Our inquiry with project leaders found two common themes for the assumptions they have about their volunteers' skill proficiency.

Assumption #1: Volunteers may not need training in order to perform the necessary skills to participate at the start.

Interviews indicated a range of perspectives in relation to how project leaders think about the skills volunteers do and do not possess and their proficiency at these skills when entering the project. All ten leaders made statements relating to this theme. Four out of the ten project leaders reported that their projects required no training before volunteers could start participating. Some project leaders assumed that volunteers understand the protocol (based on the provided information) and can perform the basic skills necessary for participation such as measurement. As one project leader explained:

> *Because it's such a voluntary program, people opt in because of a connection to content [i.e., they are already interested in the topic]...When we are expecting people to opt in from the content side, perhaps we take some shortcuts [with our training], which may be problematic.*

Thus, they provide training only on content or on more complex skills such as navigating to a field site or identifying a species. Still others provided training only on content knowledge or provided no training at all. The spectrum of perspectives on this assumption is demonstrated through quotes from two project leaders:

We don't do formal training.

When people first sign up … we mail printed materials with written detailed instructions and examples of how to count… Tips and tricks for how to identify [species] is in a mailed handbook that we give them. The volunteers go through an [in-person] training, learning about the protocol that we use and [species] identification, and then we do a practice survey. There is also online training courses if there is not a local chapter, or if people would want to refresh their skills or familiarity or practice anything.

Assumption #2: Practice makes perfect.

Four project leaders also assume that volunteer proficiency will improve over time and with experience in the project. For example, one project leader stated, "I do feel like you could think of [species] identification as a specialized skill, maybe, and that gets better over time, I would assume," while another one stated, "What we think.. [is by] doing [this] through time, you get better at noticing the [animal]."

Project leaders often justified their assumptions by explaining that their projects provide written documents that volunteers could reference if they had questions. For example, one project leader remarked, "I definitely went in with the assumption that people were using the help resources way more than they actually are." Another project leader explained this more fully:

…my assumption would be that the longer you participated, the more time you spent looking at the identification guides, the more time you've spent possibly doing research elsewhere to try to figure out if you're submitting a photo or your set of photos. And [if] you're trying to figure out which of the species it is, you might start with our identification guides. We have some resources on our site, but some of our participants also start Googling and go looking for other information and start trying to build their own knowledge. And so I would expect it would be from that research that they're doing on our site, on other sites, where they're being exposed to that language and exposed to how others are describing it. That would be my hunch.

One project leader revealed that they had previously assumed that practice makes perfect, but by conducting an evaluation prior to our work together, they had discovered the active role that project leaders need to play in training volunteers to be proficient at inquiry skills.

…we had done a study of our volunteers at how accurate they were in identifying these animals. The assumption that we had was that they would improve over time, having seen so many images. And what we found was that they weren't…really improving in their skills for identifying animals over time [when they] weren't getting feedback from us. So, that's why that feature was implemented…we felt that giving them feedback would improve their performance.

VOLUNTEERS' SCIENTIFIC OBSERVATION SKILLS

This study found that the majority of CS volunteers on the seven study projects had the necessary skill proficiency to collect accurate scientific observations. Specifically, 72% of volunteer responses accurately recorded standard observations, 81% of volunteers among the three projects accurately identified species, and 65% of volunteers could accurately notice at least one feature that is considered relevant to identifying the organism. Many of the volunteers could notice two or more features (Figure 1).

Results also indicated that volunteers who could correctly notice at least one relevant feature of an organism were more likely to identify the species accurately (M = 0.97, SD = 0.17) than those who could not (M = 0.59, SD = 0.49), $t(291)$ = 9.24, p < .05). The effect size for this analysis (d = 1.03) was found to exceed Cohen's (1988) convention for a large effect (d = 0.80).

We did find a small but significant difference between volunteers' ability to record different kinds of biological observations accurately. That is, volunteer observations of animals (M = 0.76, SD = 0.43) were more likely to be accurate than their observations of plants (M = .66, SD = 0.47), t(396) = 2.07, p < .05, Cohen's d = 0.22). Returning volunteers were significantly more likely to *notice relevant features* (M = 1.17, SD = 0.93) than new volunteers (M = 0.93, SD = 1.02), $t(431)$ = 2.57, p < .05, Cohen's d = 0.24).
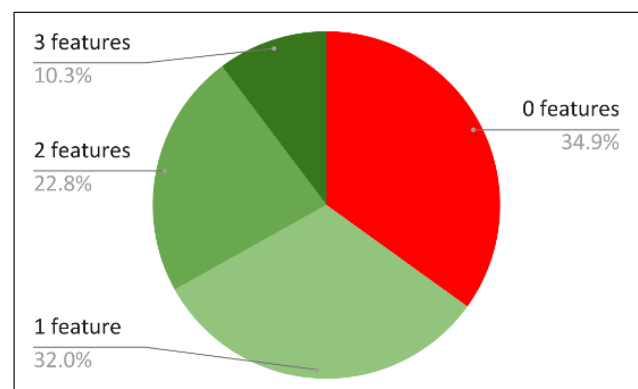

Figure 1 Percentage of volunteers who *notice relevant features*.

In addition, returning volunteers were more likely to identify an organism accurately to the species level (M = 0.87, SD = 0.33) than new volunteers (M = .71, SD = 0.45), t(293) = 2.57, p < .001, Cohen's d = 0.24).

DISCUSSION

We set out to explore the assumptions that project leaders have about their volunteers' science inquiry skill proficiency and to investigate volunteers' actual proficiency at scientific observation, a skill that is fundamental to and shared by many projects. We piloted two shared embedded assessment tools focused on dimensions of scientific observation in natural settings, *notice relevant features for taxonomic identification* and *record standard observations*, to answer questions about the extent to which volunteers can perform the skills and about differences in skill proficiency based on volunteer experience and data collection procedures.

While our previous work identified organizational barriers to evaluation (Stylinski et al. 2020), this study is novel in that it identifies organizational assumptions that function as conceptual barriers to measuring skills. First, some CS projects assume that volunteers come to a project with the necessary skills to participate (without needing training), and second, they assume that volunteers improve in those skills over time through continued participation. These assumptions could influence the way in which CS projects ask for volunteer involvement. For instance, the assumption that volunteers come to a project with the needed skills could mean that projects do not find it necessary to train volunteers, which could lead to volunteers making mistakes and not collecting data accurately. These types of conceptual barriers may stand in the way of CS project efforts to assess volunteer skill.

Our research did indicate that, overall, the majority of volunteers are proficient in skill dimensions measured in this study: *notice relevant features for taxonomic identification* and *record standard observations*. The percent accuracy rates reported in this study (between 65% and 81%) are similar to acceptable success rates reported in data validation studies (65–85% and 72%) (Crall et al. 2011; Tøttrup et al. 2021, respectively). This provides empirical evidence to support the assumption held by some project leaders that their volunteers have the necessary skills to participate in the CS project.

However, the findings also suggest there could be nuances in volunteer skill proficiency based on data collection procedures and the skill assessed. In this study, volunteers were more likely to accurately *record standard observations* of animals than plants, and volunteers who were more accurate at the skill *notice relevant features* were also more likely to be accurate at the complex skill of *taxonomic identification*. These findings suggest the importance of assessing volunteer skills so that project leaders understand the training needs of their volunteers.

Not surprisingly, returning volunteers were more accurate in their observations than new volunteers. However, this was a cross-sectional study that collected data from many individuals at one point in time, and thus does not provide any evidence about whether returning volunteers got better over time; another possible explanation for this finding is that people who are better at the skill tend to stay involved in the project for longer. Assuming volunteers will increase proficiency at a skill over time without additional support from the CS project may influence how CS projects design their onboarding and training of volunteers, decreasing the likelihood of continuing education and corrective feedback. This is at odds with the recommended best practices (National Academies of Sciences, Engineering, and Medicine 2018), which advocate for ubiquitous learning design considerations that include building learning supports, such as training and frequent opportunities for feedback, into the project. That is because, while practice is integral in the process of learning a new skill, practice alone is not enough. For example, feedback coupled with corrective action and/or reinforcement are additional steps commonly recognized in the behavior change literature as necessary for learning (Heimlich and Ardoin 2008).

CONCLUSIONS

Our study contributes to the citizen science field in three fundamental ways. One, it documents assumptions held by some project leaders that are serving as conceptual barriers to implementing assessments of volunteers' skills. Two, it adds further evidence to the credibility of volunteer-collected data, while, three, demonstrating the value of cross-project analyses using a shared assessment tool.

As previously mentioned, the assumptions held by some project leaders could contribute to the limited efforts at skill assessment by CS projects and, furthermore, to the gap between intended and assessed skill-based outcomes in CS projects. Resolving these types of conceptual barriers is a first and crucial step for projects to implement assessments of volunteers' skills.

The higher accuracy rates of returning volunteers than of new volunteers suggests future research into reasons for the difference (i.e., practice, less skilled volunteers dropping out after first year). In addition, our study points

to the importance of data collected about volunteers as critical for future investigations. A limitation of our study was that we could not conduct an analysis with volunteer training as a variable because only one project collected that information.

There are many ways that this focused examination of skills can be applied to different stages of CS project design and operation, as outlined by Davis et al. (2022). For instance, skill assessment can inform the type of data a CS project collects, keeps, and discards. It can also inform volunteer recruitment strategies, training topics, and training delivery methods. Identifying skill gaps and updating training or program protocols to fill those gaps can improve data quality proactively. Assessing skills during ongoing participation supports targeted and useful feedback to participants, which can be valuable for volunteer retention and continuous performance improvement (Van der Wal et al. 2016). In team-based projects, a skill assessment could be utilized to form data collection groups, distributing volunteers proficient in the necessary skill across all groups. When skill assessments highlight regularly occurring issues that might otherwise go undetected, data validation processes can include steps to address those issues specifically. Skill assessments can also be used in analyses; this could include developing models that weight data based on assessed skill levels (Hines et al. 2015; Kelling et al. 2015). Beyond the implications for individual projects, volunteers who demonstrate proficiency in a certain skill for one project could be considered pre-qualified for another, thereby streamlining the training process for organizations that are often resource-strapped. Additionally, this facilitated process could be utilized by research projects that use badging and micro-credentials (Fischer, Oppl and Stabauer 2022; Fishman, Teasley, and Cederquist 2018).

The need to assess skills and the challenges associated with measuring them are not unique to CS. Indeed, both formal and informal science education have seen calls encouraging researchers and evaluators to begin using performance as a key metric of skill (Bell et al. 2009; Fenichel and Schweingruber 2010; National Research Council 2014). Skill assessments like those reported in our study are poised to make significant contributions to science education at large.

DATA ACCESSIBILITY STATEMENT

The data used in the research project has not been made available, in accordance with our Internal Review Board's determination about the best way to protect confidentiality.

SUPPLEMENTARY FILE

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

- **Supplemental File 1.** Appendices A–C. DOI: https://doi.org/10.5334/cstp.628.s1

ETHICS AND CONSENT

This research was approved by the University of Maryland Institutional Review Board (IRB #1072528). Informed consent was obtained from all participants before being interviewed.

ACKNOWLEDGEMENTS

FUNDING INFORMATION

COMPETING INTERESTS

CD, KP and TP are on the editorial board of *Citizen Science: Theory and Practice*, which is on a voluntary basis. All other authors have no competing interests.

AUTHOR AFFILIATIONS

Veronica Del Bianco orcid.org/0000-0001-9088-9488
University of Maryland Center for Environmental Science, US
Karen Peterman orcid.org/0000-0003-4388-9412
Catalyst Consulting Group, US
Rachel Becker-Klein orcid.org/0000-0002-1456-3491
Two Roads Consulting, US

Cathlyn M. Davis orcid.org/0000-0002-0968-4336
University of Maryland Center for Environmental Science, US
Andrea Grover orcid.org/0000-0003-4082-4138
University of Nebraska Omaha, US
Tina Phillips orcid.org/0000-0002-5010-6052
Cornell University, US

REFERENCES

Becker-Klein, R, Davis, C, Phillips, T, Del Bianco, V, Grack Nelson, A and **Christian Ronning, EC.** 2023. Using a shared embedded assessment tool to understand participant skills: Processes and lessons learned. *Citizen Science: Theory and Practice*, 8(1): 20. DOI: https://doi.org/10.5334/cstp.487

Bela, G, Peltola, T, Young, JC, Balázs, B, Arpin, I, Pataki, G, Hauck, J, Kelemen, E, Kopperoinen, L, Van Herzele, A and **Keune, H.** 2016. Learning and the transformative potential of citizen science. *Conservation Biology*, 30(5): 990–999. DOI: https://doi.org/10.1111/cobi.12762

Bell, P, Lewenstein, B, Shouse, AW and **Feder, MA.** (eds.) 2009. *Learning Science in Informal Environments: People, Places, and Pursuits*. Washington, DC: National Academies Press.

Bonney, R, Cooper, CB, Dickinson, J, Kelling, S, Phillips, T, Rosenberg, KV and **Shirk, J.** 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11): 977–984. DOI: https://doi.org/10.1525/bio.2009.59.11.9

Bowser, A, Cooper, C, de Sherbinin, A, Wiggins, A, Brenton, P, Chuang, T.-R, Faustman, E, Haklay, M and **Meloche, M.** 2020. Still in Need of Norms: The State of the Data in Citizen Science. *Citizen Science: Theory and Practice*, 5(1): 18. DOI: https://doi.org/10.5334/cstp.303

Burgess, HK, DeBey, LB, Froehlich, HE, Schmidt, N, Theobald, EJ, Ettinger, AK, HilleRisLambers, J, Tewksbury, H and **Parrish, JK.** 2017. The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, 208: 113–120. DOI: https://doi.org/10.1016/j.biocon.2016.05.014

Cohen, J. 1988. Set correlation and contingency tables. *Applied Psychological Measurement*, 12(4): 425–434. DOI: https://doi.org/10.1177/014662168801200410

Crall, AW, Newman, GJ, Stohlgren, TJ, Holfelder, KA, Graham, J and **Waller, DM.** 2011. Assessing citizen science data quality: An invasive species case study. *Conservation Letters*, 4(6): 433–442. DOI: https://doi.org/10.1111/j.1755-263X.2011.00196.x

Davis, C, Del Bianco, V, Peterman, K, Grover, A, Phillips, T and **Becker-Klein, R.** 2022. Diverse and Important Ways Evaluation can Support and Advance Citizen Science. *Citizen Science: Theory and Practice*, 7(1): 30. DOI: https://doi.org/10.5334/cstp.482

Eberbach, C and **Crowley, K.** 2009. From everyday to scientific observation: How children learn to observe the biologist's world. *Review of Educational Research*, 79(1): 39–68. DOI: https://doi.org/10.3102/0034654308325899

Fenichel, M and **Schweingruber, HA.** 2010. Surrounded by science: Learning science in informal environments. *Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.

Fernandez-Gimenez, ME, Ballard, HL and **Sturtevant, VE.** 2008. Adaptive management and social learning in collaborative and community-based monitoring: a study of five community-based forestry organizations in the western USA. *Ecology and Society*, 13(2). DOI: https://doi.org/10.5751/ES-02400-130204

Fischer, T, Oppl, S and **Stabauer, M.** 2022. *Micro-credential Development: Tools, Methods and Concepts Supporting the European Approach*. Paper presented at 17th International Conference on Wirtschaftsinformatik, Nuremberg, Germany. https://www.researchgate.net/publication/356568140_Micro-Credential_Development_Tools_Methods_and_Concepts_Supporting_the_European_Approach.

Fishman, B, Teasley, S and **Cederquist, S.** 2018. Micro-credentials as evidence of college readiness: Report of an NSF workshop. https://deepblue.lib.umich.edu/handle/2027.42/143851.

Fu, AC, Kannan, A and **Shavelson, RJ.** 2019. Direct and unobtrusive measures of informal STEM education outcomes. *New Directions for Evaluation*, 2019(161): 35–57. DOI: https://doi.org/10.1002/ev.20348

Furnham, A. 1986. Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7(3): 385–400. DOI: https://doi.org/10.1016/0191-8869(86)90014-0

Haywood, BK, Parrish, JK and **Dolliver, J.** 2016. Place-based and data-rich citizen science as a precursor for conservation action. *Conservation Biology*, 30(3): 476–486. DOI: https://doi.org/10.1111/cobi.12702

Heimlich, JE and **Ardoin, NM.** 2008. Understanding behavior to understand behavior change: A literature review. *Environmental Education Research*, 14(3): 215–237. DOI: https://doi.org/10.1080/13504620802148881

Hines, G, Swanson, A, Kosmala, M and **Lintott, C.** 2015. Aggregating user input in ecology citizen science projects. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(2): 3975–3980). DOI: https://doi.org/10.1609/aaai.v29i2.19057

Kelling, S, Johnston, A, Hochachka, WM, Iliff, M, Fink, D, Gerbracht, J, Lagoze, C, La Sorte, FA, Moore, T, Wiggins, A and **Wong, WK.** 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? *PloS One*, 10(10): e0139600. DOI: https://doi.org/10.1371/journal.pone.0139600

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 Research Council. 2012 *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* National Academies Press. https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts.

National Research Council. 2014. *Developing assessments for the next generation science standards.*

Peter, M, Diekötter, T and **Kremer, K.** 2019. Participant outcomes of biodiversity citizen science projects: A systematic literature review. *Sustainability,* 11(10): 2780. DOI: https://doi.org/10.3390/su11102780

Peter, M, Diekötter, T, Kremer K and **Höffler, T.** 2021. Citizen science project characteristics: Connection to participants' gains in knowledge and skills. *PLoS ONE,* 16(7): e0253692. DOI: https://doi.org/10.1371/journal.pone.0253692

Peterman, K, Del Bianco, V, Grover, A, Davis, C and **Rosser, H.** 2022. Hiding in Plain Sight: Secondary Analysis of Data Records as a Method for Learning about Citizen Science Projects and Volunteers' Skills. *Citizen Science: Theory and Practice,* 7(1): 35. DOI: https://doi.org/10.5334/cstp.476

Phillips, T, Porticella, N, Constas, 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

Richards, KAR and **Hemphill, MA.** 2018. 'A practical guide to collaborative qualitative data analysis'. *Journal of Teaching in Physical Education,* 37(2): 225–231. DOI: https://doi.org/10.1123/jtpe.2017-0084

Shirk, JL, Ballard, HL, Wilderman, CC, Phillips, T, Wiggins, A, Jordan, R, McCallie, E, Minarchek, M, Lewenstein, BV, Krasny, ME and **Bonney, R.** 2012. Public participation in scientific research: a framework for deliberate design. *Ecology and society,* 17(2). https://www.jstor.org/stable/26269051. DOI: https://doi.org/10.5751/ES-04705-170229

Stylinski, CD, Peterman, K, Phillips, T, Linhart, J and **Becker-Klein, R.** 2020. Assessing science inquiry skills of citizen science volunteers: A snapshot of the field. *International Journal of Science Education, Part B,* 10(1): 77–92. DOI: https://doi.org/10.1080/21548455.2020.1719288

Tøttrup, AP, Svenningsen, L, Rytter, M, Lillemark, MR, Møller, P and **Knudsen, SW.** 2021. Citizens in the lab: performance and validation of eDNA results. *Citizen Science: Theory and Practice,* 6(1): 35. DOI: https://doi.org/10.5334/cstp.382

van der Wal, R, Sharma, N, Mellish, C, Robinson, A and **Siddharthan, A.** 2016. The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology,* 30(3): 550–561. DOI: https://doi.org/10.1111/cobi.12705

]u[