

A Rubric to Evaluate Citizen-Science Programs for Long-Term Ecological Monitoring

CATHERINE A. TREDICK, REBECCA L. LEWISON, DOUGLAS H. DEUTSCHMAN, TIMOTHY ANN HUNT, KAREN L. GORDON, AND PHOENIX VON HENDY

Citizen-science (CS) programs provide a cost-effective way to collect monitoring data over large temporal and spatial scales. Despite the recent proliferation of these programs, some in the conservation and management community remain skeptical about the quality of information generated, in part because of the lack of a rigorous framework for program evaluation. Drawing from the CS literature, we developed a structured rubric to guide the evaluation of CS programs. We test the utility of the rubric by conducting an internal and external review of a case-study CS program. The case study demonstrates the importance of the evaluation process and the effectiveness of the rubric to identify program elements that needed improvement. Our results support the assertion that program evaluation using a structured rubric can help CS programs meet their objectives, promote CS data usage in conservation and management, and maximize CS return on investment.

Keywords: citizen science, communication planning, data management, evaluation rubric, program evaluation

Citizen-science (CS) programs are expanding and increasingly being used across the United States and other countries (Bell et al. 2007, Bonney et al. 2009, Crall et al. 2010, Dickinson et al. 2010, Burton 2012, Matteson et al. 2012, Tulloch et al. 2013, Bonney et al. 2014, Donnelly et al. 2014, McKinley et al. 2016). We broadly define CS as projects in which members of the public collect, categorize, transcribe, or analyze scientific data (Bonney et al. 2014). Evidence of the growing interest and energy directed toward CS can be found on the websites of several societies and organizations across the globe (*citsci.org*, developed through the Natural Resources Ecology Lab at Colorado State University; the Citizen Science Association, *citizenscience.org*; the Citizen Science Alliance, *citizensciencealliance.org*; the Australian Citizen Science Network, *citizenscience.org.au*; the European Citizen Science Association, *www.citizen-science.net*). With the proliferation of CS groups and organizations has also come a recognition of the social importance and power of local participatory efforts in resource monitoring that extends beyond data collection (Constantino et al. 2012, Funder et al. 2013, Predavec et al. 2016, Schmiedel et al. 2016), including improving scientific literacy in communities, increasing public support for and commitment to conservation and stewardship, improving public participation in the planning and management of local ecosystems, and collecting usable long-term data

at multiple scales (Cooper et al. 2007, Danielsen et al. 2007, Dickinson et al. 2010, Conrad CC and Hilchey 2011, Constantino et al. 2012, Jordan et al. 2012a, Bonney et al. 2014, Chandler et al. 2016).

This latter objective, long-term data collection, is evident in the many CS programs that have been developed to monitor abundance and richness trends in a variety of taxa (e.g., anurans, De Solla et al. 2005; birds, Bonter and Harvey 2008, Jiguet et al. 2012; butterflies, Matteson et al. 2012; and macroalgae and invertebrates in the rocky intertidal, Cox et al. 2012; see also table 1 in Chase and Levine 2016). With long-term data, CS programs can be used as early warning systems in water-quality monitoring (e.g., Mullen and Allison 1999), detecting invasive plant species (Crall et al. 2010, Jordan et al. 2012b), and monitoring fisheries and other resource-extraction activities (e.g., Sultana and Abeyasekera 2008). CS programs are also increasingly used as an efficient and cost-effective way to amass large data sets for monitoring ecological patterns and processes across large spatial and temporal scales (Bonney et al. 2009, Crall et al. 2010, Dickinson et al. 2010, Hochachka et al. 2012, Hunter et al. 2013, Tulloch et al. 2013, Forrester et al. 2015). Some of the most well-known CS programs are designed with this objective in mind, such as the US and UK Breeding Bird Survey (BBS), the Audubon Christmas Bird Count (CBC), the UK Wetland Birds Survey (WeBS), and the UK

Table 1. Criteria for evaluating a citizen-science program.

Element	Subelement	Criteria
1) Stakeholder collaboration and program resources	1a) Stakeholders	Have key stakeholders been included in key steps of program development and implementation? Is there a structured scheme for linking program participants? Does the team include scientists, technologists, and participants? Is there mutual trust among program participants?
	1b) Resources	Have resources available for the program (e.g., money, expertise, and participants) been assessed? Is there a long-term commitment for funding? Are there adequate staff with appropriate training?
	1c) Volunteers	Is there infrastructure for the recruitment and retention of volunteers? Are there opportunities for volunteers to see the outcomes and outputs from their work?
2) Goals and objectives	2a) Goals	Is there a clearly defined goal of the program? Is the goal providing information that will influence conservation or management outcomes? Does the goal lend itself to recruitment and retention of volunteers?
	2b) Objectives	Are the objectives aligned with the overall goal? Are the objectives SMART (specific, measurable, attainable, relevant, and timely)? Are the objectives easy to explain and understand?
3) Methods: Design and implementation of monitoring	3a) Current understanding and conceptual model	Has all existing information on the system, background, and methods been compiled? Has a conceptual model of the system been created? Does the conceptual model link the goals and objectives to the information needed? Does the data collected by the program inform the model and hypothesis?
	3b) Sample and protocol design	Is the sampling protocol well designed (e.g., is scientifically sound, uses established methods, and explicitly considers power and sample size) and easy for volunteers to follow? Are there specific hypotheses that are being tested (if relevant)? Is the sampling design appropriate for the objectives? Are the response metrics relevant and sensitive to change, and can they be measured against an appropriate reference state? Have the methods and protocols been tested (pilot data)? Have appropriate analysis methods been outlined <i>a priori</i> ? Are planned statistical analyses being considered as part of monitoring program activities?
	3c) Training and managing volunteers	Are protocols easy to understand and implement and appropriate for the level of expertise of the volunteers? Is there adequate training of volunteers? Have the training materials been evaluated for clarity and effectiveness? Are the techniques evaluated and verified in the field when adopted by volunteers?
4) Data entry, storage, analysis, and synthesis	4a) Organization and management of data	Are data sets organized and well documented? Are data housed in secure storage with long-term searchable archives? Is there a coordinator to maintain the data and screen for errors? Are data property and rules of access clear? Are methods for uploading and downloading simple and clear?
	4b) Quality assurance and information integrity	Has the data-entry method been tested? Are there detailed specifications for methods, data entry, and Qa/Qc? Are the data-quality or data-assurance filters adequate? Are data-validation or -verification measures being used? Are there regular review and quality checks for the data and database?
	4c) Data analysis and interpretation	Does the analysis provide intended information about program goals and objectives? Is the right amount of data being collected (not too much or too little)? Are the data being collected at the correct scale(s)? Are the current methods appropriate based on analysis of current monitoring data?
5) Reporting and dissemination	5a) Communication planning	Is there a clear commitment for public dissemination to provide educational value and facilitate community involvement? Does the communication plan identify the audience and therefore the best medium for communication? Is there a comprehensive communication strategy for disseminating results? Is the communication plan designed to feedback to conservation and management?
	5b) Outreach implementation and reporting	Is reporting regular, with results that are available quickly? Are results being presented at the appropriate level of detail (scientific rigor)? Are results being disseminated so that conclusions translate into action? Does the communication include press releases, scientific publications, networking opportunities, and educational outputs? Does communication reach and engage the public (e.g., an interactive website)?
6) Outcome assessment and program review	6a) Evaluating outcomes: Science, learning, and engagement	Are scientific outcomes being assessed and measured? Is the monitoring program informing conservation outcomes or outputs? Are the data being used in decision-making? Is the information contributing to the peer-reviewed scientific literature (if appropriate)? Is the program contributing to learning and engagement?
	6b) Program review: Self study and/or external review	Do the participants reflect (periodically) on the programs strengths and weaknesses? Is there a mechanism for evaluation and feedback throughout the implementation of the program? Are the participants and stakeholders open to external peer review? Are reviews planned to be formative (less formal and part of program formation or development) or summative (periodic external reviews)?

Note: Elements, subelements and associated criteria were created based on a systematic review of the CS literature. Papers were included if they explicitly critiqued CS or monitoring programs (e.g., table 2 in Donnelly et al. 2014) or if they described strengths (desirable properties, keys to success) and weaknesses (challenges) of CS programs (e.g., table 2 in Conrad and Hilchey 2011 or figure 1 in Devictor et al. 2010). Each of the elements, subelements, and criteria is based on information from at least 10 references. A full list of references identified and used is available in supplemental document S1.

Butterfly Monitoring scheme (UKBMS; Atkinson et al. 2006, Dickinson et al. 2010, Brereton et al. 2011, Conrad CC and Hilchey 2011). A recent review highlighted the substantial monitoring data that CS programs have already collected, particularly for birds, butterflies, and plants (Chandler et al. 2016). Although the degree of citizen involvement varies widely across these programs (Danielsen et al. 2009), they have been found to be well suited for monitoring larger areas over longer timescales given funding gaps at agencies and academia and the relatively short time horizons for most research projects (Whitelaw et al. 2003, Crall et al. 2010, Dickinson et al. 2010, Tulloch et al. 2013, Predavec et al. 2016).

For CS to be widely recognized as a valid means to collect long-term monitoring data, it is essential that programs be able to demonstrate that their data can support conservation and management decisions. Despite recent quantitative analyses that reveal strong concurrence between citizen-collected and scientist-collected data (Danielsen et al. 2005, Danielsen et al. 2014, Dolrenry et al. 2016, Predavec et al. 2016) and that show that data quality from CS programs is robust (Schmeller et al. 2008), many continue to question the validity of data collected by CS because of concerns about data collection and management, including the use of inappropriate methods and/or inconsistent sampling (Danielsen et al. 2005, Cooper et al. 2007, Silvertown 2009, Crall et al. 2010, Dickinson et al. 2010, Burton 2012, Tulloch et al. 2013, Bonney et al. 2014, Cooper et al. 2014, Lukyanenko et al. 2016); the nonuniform competency and retention of volunteers (Kremen et al. 2011, Cox et al. 2012, Jordan et al. 2012a, Matteson et al. 2012, Moyer-Horner et al. 2012); and inadequate data management, including data entry, quality control, and timely analysis of complex and often heterogeneous data (Danielsen et al. 2005, Dickinson et al. 2010, Newman et al. 2011, Bonter and Cooper 2012, Hunter et al. 2013, Donnelly et al. 2014). Periodic program evaluation—that is, an internal and external review or performance audit—of CS programs can play a vital role in identifying and overcoming these potential shortcomings and, as a result, can help maximize the return on investment of a CS program (Crall et al. 2010, Dickinson et al. 2010, Newman et al. 2011, Shirk et al. 2012, Tulloch et al. 2013). We use the term *return on investment* generally, referring not to a formal cost-benefit analysis (e.g., Tulloch et al. 2013) but rather to more general, multifaceted benefits that can be gained by examining and refining elements of CS monitoring programs (sensu Possingham et al. 2012).

A number of guidelines have been developed to support the establishment of a well-designed CS monitoring program (Cooper et al. 2007, Conrad CT and Daoust 2008, Bonney et al. 2009, Shirk et al. 2012, Donnelly et al. 2014, Pocock et al. 2014, Shirk and Bonney 2015). Although there is no one-size-fits-all approach, these guidelines identify a core set of principles that contribute to successful CS programs. Despite these general establishment guidelines, there is no functional program-evaluation framework to guide a CS

program review. To address this gap in program-evaluation structure and process, we developed a formal assessment rubric for CS program evaluation on the basis of a comprehensive review of the published and available CS literature.

The development of a formal CS assessment rubric fills an important gap in the CS landscape, providing these programs with a critical tool and a means to formally evaluate CS program elements. The rubric supports a tiered approach to program evaluation that can be deployed to meet the specific needs and goals of individual CS programs for specific program elements (e.g., various levels of stakeholder involvement or outreach) or can be expanded to include a comprehensive review that includes self-evaluation, an external review, and thorough data analysis. Use of a formalized program-evaluation rubric serves to improve the utility and the applicability of CS data, which in turn directly focuses and enhances the scope, application, and return on investment of long-term monitoring and other CS programs.

Systematic literature review

Systematic reviews are becoming widespread in ecology and conservation as a way to synthesize current understanding in a formal and transparent way (Pullin and Stewart 2006, Cook et al. 2013, Lorkie 2014, Doerr et al. 2015). Although a formal meta-analysis was not appropriate for this topic, we developed a systematic review of the literature following the current practices outlined in Doerr and colleagues (2015). The key steps of this review include developing an explicit method to identify potential papers, refining and reducing the initial list of papers to a smaller list of papers that provided explicit information on our topic of interest, and extracting information from these papers in a way that is replicable (Lorkie 2014, Doerr et al. 2015).

Our systematic review process was conducted in three stages. First, we started with a pilot phase in which we validated our search strategy and choice of keywords within Web of Science. We included papers that used the term *citizen science* in the title and had keywords of *conservation*, *assessment*, or *monitoring*. We then broadened the search to all papers that used the term *citizen science*, *citizen scientist*, or *local participatory* in the title. We then used the intersection between *citizen science* or *volunteer* in the topic and additional keywords including *conservation*, *assessment*, or *monitoring*. We explicitly excluded those with *citizen science* in the title so that there was no overlap with the first two searches. The three searches yielded 703 references that were then reviewed for inclusion in the rubric (supplemental document S1). In the second stage of our review, we refined and reduced this list of references on the basis of keywords and a review of abstracts. Finally, in the third stage, we developed a template to synthesize and extract relevant information for each element of the rubric from the text of the identified articles.

Because CS programs are often designed to provide long-term monitoring data, our rubric also draws from the resource monitoring literature to capture best practices in

this area (e.g., Schroeder 2009, Sergeant et al. 2012). For example, we discuss the need for clear and concise goals and objectives, where a *goal* is a broad, concise visionary statement that defines the intended purpose of a monitoring program (Adamcik et al. 2004, Tear et al. 2005) and *objectives* conform to SMART criteria (specific, measurable, achievable, results oriented, and time fixed; Adamcik et al. 2004, Schroeder 2009). However, our literature review was not meant to concurrently nor comprehensively review the long-term monitoring literature. Rather, we focus our attention on the citizen-science literature while highlighting some of the seminal papers on long-term monitoring.

Rubric development

From the literature review, we identified, organized, and synthesized more than 224 concepts and ideas from the relevant papers into six elements that constitute the final rubric: (1) stakeholder collaboration and program resources; (2) goals and objectives; (3) design and implementation of monitoring; (4) data entry, storage, analysis, and synthesis; (5) reporting and dissemination; and (6) outcome assessment and program review. These elements were further subdivided into subelements to provide the necessary detail to represent the current state of CS knowledge from the literature (table 1). On the basis of the literature, we also created performance levels for each element of the rubric (table 2). The structured framework of the rubric, with elements and specified subelements, presents a step-by-step approach for evaluating the many dimensions of a CS program.

Rubric deployment: Internal and external review process

We piloted the rubric and the review process using an established CS monitoring program in San Diego County, California, USA, the San Diego Tracking Team (SDTT), as a case study. Established in 2001, the SDTT monitors mammals across a large-scale network of protected areas and preserves in San Diego County and provides the only consistent, long-term, multispecies monitoring data set in this region. As such, SDTT serves as an excellent case study for evaluating CS programs aimed at providing long-term monitoring data. Our test of the rubric included an internal review by the SDTT program leaders and an external review, including comprehensive data analysis, that we conducted. Prior to external review, we solicited basic program information from SDTT leadership. A questionnaire was developed on the basis of the rubric to better solicit specific answers as to program functioning within each element of the rubric (supplemental document S2). Both the internal and external review teams completed the questionnaire independently. Because no comprehensive data analysis had been conducted for the SDTT program, we also conducted data analyses as part of the external review to evaluate protocols, sampling design, and data collected. To evaluate the program using the rubric, we compared the narrative responses and numerical scores using a

1–10 rating scale (1, *needs considerable improvement*; 10, *no improvements required*) for each element of the questionnaire between the internal and external review teams and identified common or divergent scores, responses, or themes.

Rubric structure

Element 1 of the rubric evaluates stakeholder collaboration, resources, and volunteer participation. This element captures the foundation or backbone of a CS program. Stakeholder engagement has been identified as critical to successful CS programs. Successful CS programs should encompass and engage a multiskilled team of stakeholders, including local volunteers and citizens, academic and government scientists, conservation and management partners, statisticians, technologists, educators, and evaluators (Cooper et al. 2007, Greenwood 2007, Bonney et al. 2009, Mackechnie et al. 2011, Gallo and Waitt 2011, Jordan et al. 2012a, Donnelly et al. 2014). Numerous authors also emphasized the importance of a formalized structure for linking and engaging these stakeholders via “in-person” or alternate communication methods that facilitate relationship building (Devictor et al. 2010, Gallo and Waitt 2011, Connors et al. 2012, Jordan et al. 2012a). Several authors suggest the importance of forming a scientific advisory board to oversee protocol development, data analysis, and report writing (Tulloch and Szabo 2012, Tulloch et al. 2013, Riesch and Potter 2014) and explicitly specifying the different roles and responsibilities of the various stakeholders to avoid unnecessary overlap and confusion (Sergeant et al. 2012).

Resources are, of course, the lifeblood of any CS program, which includes operating resources, personnel resources, as well as the institutional resources needed to provide long-term support to host and run the program. As with the majority of long-term monitoring programs, resources for CS programs are often limited (Powell and Colin 2008, Crall et al. 2010, Tulloch et al. 2013, Westgate et al. 2013). Innovative use of technology (e.g., cyberinfrastructure) can help minimize the “burn rate” of resources (Newman et al. 2012), but the need for grants or trusts that allow for cross-disciplinary and sustainable work is paramount (Sharpe and Conrad 2006, Devictor et al. 2010, Conrad CC and Hilchey 2011, Gallo and Waitt 2011, Bonney et al. 2014, Crain et al. 2014, Aceves-Bueno et al. 2015). However, these resources alone may not be predictors of program quality (Nerbonne and Nelson 2008). Volunteer recruitment and retention is an equally important aspect of CS programs because volunteers are vital to the survival and sustainability of any program (Bell et al. 2008, Conrad CC and Hilchey 2011, Tulloch and Szabo 2012, Tulloch et al. 2013, Beirne and Lambin 2013, Havens and Henderson 2013). Scientists have found several elements that positively influence individual decisions to participate and remain in these types of programs, including connecting veteran and novice participants (Beirne and Lambin 2013), aligning data collection with volunteer

Table 2. Evaluation rubric for a citizen-science program.

Element	Aspect of program being reviewed	Poor	Fair	Good	Excellent
(1) Stakeholder collaboration and program resources	Considers how the program identifies, links, and engages stakeholders, trust level with stakeholders, funding sources and security, and volunteer recruitment and stability.	Little stakeholder engagement or connection. Little diversity in expertise of participants. Low level of trust with stakeholders. Little or no financial stability or planning. Poor recruitment and retention. Little staff stability.	Limited stakeholder engagement or connection. Some diversity of expertise among participants. Some established trust. Limited resources identified. Limited support of volunteer recruitment and retention. Some staff stability.	Clear stakeholder engagement and connection. Moderate trust but potentially inconsistent interaction with stakeholders. Good trust with stakeholders. Established financial planning. Good staff recruitment, retention, and stability.	A high level of consistent engagement or connection with stakeholders. A high level of expert, diverse participants. Excellent trust with stakeholders. Current and future funding sources are identified and secured. A high level of recruitment and retention. Long-term staff stability.
(2) Goals and objectives	Evaluates how well the goals and objectives have been articulated and aligned with program activities and the data collected.	The goal of the program is poorly defined or articulated, with little or no link to how the program data will influence conservation or management outcomes. The objectives are not aligned with the goal, do not meet SMART criteria, and are diffuse and convoluted.	The goal of the program is defined, but the link between program data and conservation or management outcomes is unclear. The objectives are articulated but may still be unclear and not developed to SMART criteria.	The goal of the program is defined, and there is a clear link between program data and conservation or management outcomes. The objectives are clear and fairly detailed but are not fully aligned to SMART criteria.	The goal of the program is clearly defined, and the link between the program data and conservation and management outcomes is clearly articulated. The objectives meet all SMART criteria and are prioritized on the basis of their contribution to the program mission.
(3) Methods: Design and implementation of monitoring	Assesses information used to support the sampling design (i.e., conceptual models, existing data, and knowledge). Evaluates sampling protocols and response metrics: Are they aligned with program goals and objectives, best-practices, and established methods? Are statistical or quantitative analyses part of monitoring activities? How are volunteers evaluated?	No conceptual model of the study system has been developed, and little existing knowledge has been synthesized or integrated. The goals and objectives of the program do not reflect current data or knowledge gaps. Neither sampling protocols nor monitored response metrics are based on established methods and best practices. Sampling protocols do not align with objectives. No consideration is given to sample size, statistical power, or statistical analyses. No standardized participant training or vetting process.	A conceptual model of the study system has been developed, but it draws little from existing knowledge. The goals and objectives of the program are loosely connected to current data or knowledge gaps. Sampling protocols and monitored response metrics are generally based on established methods and best practices. Sampling protocols address some objectives. Some consideration of sample size, statistical power, or statistical analyses. Limited standardized participant training or vetting process.	A conceptual model has been constructed that builds directly from existing data and knowledge. The goals of the program align well with this model, but specific objectives may be less directly linked to the model. Sampling protocols and monitored response metrics are based on established methods and best practices. Sampling protocols address most objectives. Sample size and statistical power have been formally evaluated. Statistical analyses have been considered. Moderate standardized participant training or vetting process.	A well-defined, knowledge-informed conceptual model has been developed. The goals and objectives of the program directly address uncertain or poorly understood model elements. Sampling protocols and monitored response metrics directly reflect established methods and best practices. Sampling protocols directly address objectives. Sample size and statistical power have been formally evaluated. A plan for statistical analyses is in place. Standardized and detailed participant training or vetting process.
(4) Data entry, storage, analysis, and synthesis	Considers data quality assurance and quality controls, including timely data entry, organization, metadata, data personnel, and data validation. Evaluates the rigor of statistical or quantitative analyses and assesses how data are used to improve or change sampling protocols or other elements of data collection (e.g., frequency and timing).	Data are poorly organized and managed. Open access to database and data entry. Data quality not assessed. No data analyses conducted. No data interpretation. No feedback loop by which data analyses can inform data-collection protocols.	Data are organized and managed by multiple participants. Some database access restrictions. Minimal data quality assessments. Some data analysis occurs but may be limited to simple statistical summaries. Limited data interpretation. Limited feedback to data-collection protocols.	Data are organized and managed by focal personnel. Database restrictions are in place. Data quality assessments are in place. Statistical analyses are adequate to address most of the key objectives. Some data interpretation occurs periodically. Data review is used to inform data-collection protocols.	Highly organized and managed data entry. The access and management process and data quality are regularly assessed using standardized methods. Database management includes comprehensive QaQc procedures and metadata. Rigorous statistical analyses are conducted relevant to monitored metrics and objects. Comprehensive data interpretation that is frequent and directly used to inform data-collection protocols.

Table 2. Continued.

Element	Aspect of program being reviewed	Poor	Fair	Good	Excellent
(5) Reporting and dissemination	Considers overall communication strategy and the mode and frequency of reporting technical and nontechnical results. Assesses the efficacy of the communication strategy and describes stakeholder uptake.	No communication plan in place. Little or no dissemination of program outcomes or findings to stakeholders. Poor commitment to communication to the general public or the scientific community.	A communication plan is in place. Reports or materials produced that describe outcomes or findings are infrequent or cursory in nature. Some communication with the general public and the scientific community.	A well-defined communication plan is in place. Some aspects of the program are described in materials that are prepared and disseminated fairly regularly with the general public and the scientific community.	A well-defined communication plan is in place. Comprehensive and accessible reports and materials are produced regularly and shared with stakeholders, the general public, and the scientific community.
(6) Outcome evaluation and program review	Determines how the outcomes of the program are assessed and evaluated on the basis of on scientific products, use of data by decision-makers, learning and engagement outcomes of participants, and the frequency and extent of formal and informal internal and external review.	Little or no evidence of contribution to ongoing research or conservation or management decisions. No opportunity to solicit or collect stakeholder feedback on strengths or weaknesses. No formal or informal evaluation process in place or conducted.	Some evidence of contribution to ongoing research or conservation or management decisions. Some opportunities to solicit or collect stakeholder feedback on strengths or weaknesses. Limited periodic review of aspects of the program but no comprehensive review.	Clear evidence of contribution to ongoing research or conservation or management decisions. Multiple opportunities to solicit or collect stakeholder feedback on strengths or weaknesses. Periodic internal review of program, with results used to improve the program.	Clear and direct evidence of contribution to ongoing research or conservation or management decisions. Established and periodic opportunities to solicit or collect stakeholder feedback on strengths or weaknesses. Established protocol for both internal and external review of program, with an established feedback mechanism by which results from the review are used to improve the program.

Note: The evaluation rubric for six elements of a citizen-science program based on a systematic review of the citizen-science (CS) literature. Each element includes four levels of performance (*poor*, *fair*, *good*, and *excellent*). The rubric was tested using a long-term CS monitoring program in California. Performance levels for the case-study program, generated from a comprehensive internal and external review process using the rubric, are highlighted in bold. When there were differences in evaluation results or divergent results within an element, we expanded the review to include scores for relevant subelements. Details of the rubric development and review process are included in the main text. A full list of elements, sub-elements, and associated criteria are listed in Table 1.

expertise (Bonney et al. 2009), helping participants understand the big picture of the program, and facilitating the presentation of results to policymakers (Freitag and Pfeffer 2013, Havens and Henderson 2013).

Element 2 of the rubric focuses on setting goals and objectives. The success of any long-term monitoring program, whether involving CS or not, is contingent on the development of clear and concise goals and objectives (Yoccoz et al. 2001, Adamcik et al. 2004, Conrad CT and Daoust 2008, Schroeder 2009, Sergeant et al. 2012, Donnelly et al. 2014). Typically, a statement of goals and objectives is accompanied by a short summary of the rationale for these, including appropriate literature for these targets that describes the hypotheses or context for the activities including how the data collected will add to the current body of knowledge or influence conservation or policy at a broader level (Schroeder 2009, Newman et al. 2011, Shirk et al. 2012). This structure supports the collection of useful and actionable data (Aceves-Bueno et al. 2015) and links back to the first element of the rubric by ensuring efficient allocation of resources as well as recruitment and retention of volunteers (Conrad and Daost 2008, Devictor et al. 2010). Stated objectives must be S.M.A.R.T (*sensu* Schroeder 2009), and clearly worded so that they can be understood by an interested, but general public. Although there is some disagreement about

the ability of a single CS program to meet multiple objectives (e.g., long-term-monitoring data collection as well as education or conservation literacy; Jordan et al. 2011, 2015, Marshall et al. 2012, Havens and Henderson 2013), the need for defined and explicit goals and objectives has been well documented.

Element 3 of the rubric evaluates the design and implementation of the monitoring program and protocol. Formalizing the current understanding of a system of interest, including information from both social and natural sciences, has been found to be important for formulating specific objectives related to the long-term ecological monitoring (Margoluis et al. 2009, Lindenmeyer and Likens 2010). Conceptual models can provide the framework for synthesizing this understanding, and can facilitate constructive communication about questions that still remain about the system among stakeholders (Reed 2008, Etienne et al. 2011). In addition, the model can be used to identify appropriate questions to ask with monitoring efforts, as well as prioritizing and selecting appropriate indicators and variables to monitor (DeBlust et al. 2012, Sergeant et al. 2012). Selected monitoring variables should be representative of the system being studied, relevant to a large range of conditions, sensitive to change, and be measurable against an appropriate reference state

when relevant (Eyre et al. 2011). Once appropriate questions and variables have been selected, creating basic, clear, standardized data-collection protocols for program participants to follow is essential if data are to be used to inform conservation and resource management decisions (Bonney et al. 2009, Donnelly et al. 2014). Rigorous consideration of sample size, spatial scale, statistical analyses (including power analyses), and error-checking methods must be balanced with the need for standardized, easy-to-follow protocols that match the capabilities and interests of program participants (Couvet et al. 2008, Bonney et al. 2009, Devictor et al. 2010, Conrad CC and Hilchey 2011, Donnelly et al. 2014). Pilot testing of protocol also should be conducted. Proper training of program participants is essential for ensuring data quality and volunteer retention (Framstad et al. 2008, Bonney et al. 2009, Donnelly et al. 2014, Phillips et al. 2014, van der Wal et al. 2016). Allowing participants to collect practice data prior to actual data collection and having program veterans accompany novice citizen scientists have been shown to enhance CS data quality (Gollan et al. 2012, Aceves-Bueno et al. 2015). Training should be continual, with regular feedback to participants to ensure consistent sampling and continuity of program participants (Beaubien and Hamann 2011, Donnelly et al. 2014, van der Wal et al. 2016).

Element 4 of the rubric involves the organization and management of data. As with all monitoring data, CS programs need rigorous data-entry protocols to ensure the integrity and quality of the data collected (Shirk et al. 2012, Tulloch et al. 2013, Donnelly et al. 2014). Once protocols have been developed and tested, data-entry coordinators should be appointed to oversee data entry, clearly organize and document collected data, and screen for errors in databases (Mackechnie et al. 2011, Tulloch and Szabo 2012, Tulloch et al. 2013). New technologies, including online data entry forms, smartphones, and filters to flag anomalous data, are providing more efficient and automated methods for collecting and entering data, thereby reducing the risk of creating data-entry errors (Bonney et al. 2009, Crall et al. 2010). New technologies also allow for easier access to large, long-term data sets (e.g., searchable, online databases), but data ownership and rules of access must be clear from the onset (Devictor et al. 2010). Adequate quality assurance and quality control ($QAQC$) protocols are also needed to ensure the integrity and quality of collected data (Gouveia et al. 2004, Conrad CC and Hilchey 2011, Donnelly et al. 2014). Data collection must undergo regular review and quality checks, which includes screening for suspect data (e.g., expert validation), observer bias and variation, and regular monitoring of performance to ensure that training and sampling design remain adequate (Cox et al. 2012). Regular analyses of CS data must be carried out to determine whether data being collected align with outlined objectives at appropriate scales (Devictor et al. 2010, Donnelly et al. 2014). This provides a means for continually updating current understanding of the system, as well as refining and revising program goals

and objectives, sampling and training protocols, and analysis methods, if necessary.

Element 5 of the rubric evaluates the reporting and dissemination in which the CS program engages. Failing to adequately and effectively share the large and growing amounts of information CS programs are able to generate represents a “missed opportunity in science and society” (Theobald et al. 2015). Developing a formal plan to communicate and disseminate data collected through CS efforts serves to maintain the strength and integrity of the program by (a) keeping the participants engaged and motivated and (b) ensuring that efforts are linked to conservation and management actions (Conrad CT and Daoust 2008, Devictor et al. 2010, Sergeant et al. 2012). In order to promote trust and continued engagement of citizens, the participants must understand the context and reasoning of what they are doing, as well as feeling that their work is being used in the decision-making process (Conrad CC and Hilchey 2011, Mackechnie et al. 2011, Constantino et al. 2012, Funder et al. 2013). This requires all stakeholders to actively engage and communicate through various methods of collaborative exchange (e.g., email, phone, and face-to-face meetings; Powell and Colin 2008) and media platforms (e.g., newsletters and social media; Marshall et al. 2012) to ensure that information reaches all appropriate audiences and stakeholders. Program results also must be disseminated to appropriate professional outlets (e.g., policymakers, scientific journals, and media outlets) via formal reports, scientific articles, websites, press releases, etc. in a timely and efficient manner to ensure that program efforts are indeed being used to inform the conservation and management decision-making process (Devictor et al. 2010, Donnelly et al. 2014).

Element 6 of the rubric considers outcome assessment and reviews the program from both a scientific, conservation, community building, and education perspective, acknowledging the multiple objectives of CS programs (Bonney et al. 2009, Freitag and Pfeffer 2013, Havens and Henderson 2013, Donnelly et al. 2014). In order to produce reliable data that can be used widely by the scientific and management communities, CS programs and products may need to participate in the peer-review process that traditional scientific studies undergo (Bonney et al. 2014) as the lack of uptake of CS-generated data has been linked to the lack of formal data analysis and review (Connors et al. 2012, Bonney et al. 2014).

Program evaluation must occur as a continuous feedback loop to ensure that protocols and outcomes consistently align to meet program goals and objectives, that programs are using best practices, and that future program activities adapt to and address lessons learned from the program (Conrad CT and Daoust 2008, Conrad CC and Hilchey 2011, Jordan 2011, Tulloch and Szabo 2012, Beirne and Lambin 2013, Havens and Henderson 2013, Tulloch et al. 2013). The evaluation process can have many elements, including summative and formative internal and external review of the program process, as well as a review of the data collected. All reviews should consider how information is flowing back to

Table 3. The internal and external review teams responded to the same questionnaire (supplemental document S2), and each team provided both narrative responses and a numerical score for each element on a scale of 1 (needs considerable improvement) to 10 (no improvements required).

Rubric element	Internal score	External score	Review summary
(1) Stakeholder collaboration and program resources	6	4	Both reviews revealed a high level of stakeholder engagement. External review identified a lack of engagement of the land management community and scientists, particularly those with statistical and sampling design expertise at program initiation. There was clear agreement on the need for long-term funding support. Short-term funding mechanisms and plans are established and working effectively. Volunteer engagement and retention is strong.
(2) Goals and objectives	8	5	The external review flagged loosely defined goals of the monitoring program, and associated objectives did not adhere to SMART criteria.
(3) Methods: Design and implementation of monitoring	4/7	4/7	Both reviews felt that the design of the monitoring program could be strengthened in terms of spatial and temporal distribution of data collected, with increased attention to statistical rigor and application. Both reviews agreed that the implementation, training, and data-collection protocols were satisfactory.
(4) Data entry, storage, analysis, and synthesis	8	8/4	Both reviews agreed that data entry QaQc procedures followed current best practices. External review identified problems with data analysis and utility because of the limitations of data collection mentioned in element 3. Improving on these design deficits would substantially improve the ability to analyze and use these data.
(5) Reporting and dissemination	8/3	7	The internal review was more critical of their current reporting and dissemination program than the external review because of the limited success of getting its message out to a broader audience, particularly land managers and management agencies. Despite this, both reviews felt that SDTT is strongly committed to reporting and disseminating their message, but their communication strategy needs to be strengthened.
(6) Outcome evaluation and program review	6	7	SDTT is clearly and strongly committed to the program review process and has undergone several informal external reviews since its inception. Both reviews revealed that more attention needs to be paid to providing data that can meet specific needs of local managers and decision-makers. Development of specific objectives, as were identified in element 2, would help in this process.

Note: For some elements, the teams provided multiple scores (#/#) to acknowledge different performance levels within an element. The areas of discordance between the internal and external review teams are shown in bold.

program participants and how data are being used by the community, stakeholders, or other entities (Newman et al. 2011, 2012). If regular and robust data analyses have been conducted by the CS program, an external review process may review the findings or results of this analysis. If rigorous or timely data analyses have not been conducted, an external review panel may also need to analyze existing data to evaluate this element of the rubric.

Rubric deployment: Case study results from the internal and external reviews

Although the internal and external reviews were conducted independently, the comments and assessments across the two reviews were fairly concordant. We have summarized the results from the internal and external responses to the program-evaluation questionnaire (supplemental document S2) in table 3. The internal and external scores for each element, as well as a brief summary and comparison of the questionnaire responses, are provided. The general performance standards for each element are listed in the evaluation rubric (table 2), with specific performance levels for the case-study program, SDTT. For some elements, the teams provided multiple scores to acknowledge different performance levels within an element.

Conclusions

Our effort to develop a comprehensive program-evaluation rubric for CS programs represents a crucial step to advance

the role CS plays in science, conservation, and resource management, because the lack of formalized program evaluation often impedes the use of CS data in these contexts (Conrad CT and Daoust 2008, Conrad CC and Hilchey 2011, Bonney et al. 2014, Chandler et al. 2016). Our rubric provides a thorough but flexible approach to evaluating the key elements of CS programs, as have been articulated by the rich CS literature, to provide formal validation and assurance that these programs are following best practices to collect data that can inform, support, and advance conservation and management decisions. This program-evaluation instrument and process also provides CS programs with a means to monitor progress and implement change as needed.

Clearly, there is no one-size-fits-all evaluation process for all CS programs that exist across the globe, even within the realm of CS programs collecting long-term resource monitoring data. Although each element of the rubric is important to the effective functioning of CS programs, there is no single, monolithic approach to how programs should be meeting program-evaluation standards. Evaluation can take place in a consistent, systematic fashion, proceeding through the rubric elements from start to finish, or may adopt a disjointed approach, starting with the more deficient elements of a program and eventually working through evaluation of all elements. Thus, our program-evaluation rubric provides a broad guide to the development and evaluation of key

elements of a CS program and can be tailored at each element to meet the specific needs of individual CS programs. Over time, this rubric and review process can be refined and streamlined as it becomes a more integral part of CS programs. With repeated use, the rubric should serve as a regular and efficient means of evaluating, refining, and improving program objectives and outputs, as well as incorporating new information and ideas to maximize utility and return on CS investment. Other rubric changes may also be warranted to support CS programs in developing countries. Although our rubric had no intended geographic bias, a significant proportion of CS literature is from developed countries. As the role of CS programs in developing communities is an emerging field of research and activity (e.g., Constantino et al. 2012, Funder et al. 2013, Danielsen et al. 2014), it will be important to consider how the rubric can best be modified or supplemented to most effectively serve these CS programs.

Our case study of the SDTT provides insight into how the evaluation process can affect CS program outcomes and effectiveness. In many instances, internal and external evaluations from this case identified similar strengths and weaknesses within each program element, suggesting that the questionnaire is an effective tool to reflect on and evaluate key elements of the program. The internal reviewers were able to identify and acknowledge where improvements could be made, as well as tout elements of program success. Most areas of discordance between the internal and external reviews related to best practices from the scientific literature that were familiar to the external but not to the internal review team. For example, the significance of SMART objectives and spatial and temporal scale and statistical design issues that are commonly discussed in the monitoring literature were identified by the external but not the internal reviewers. These differences illustrate the advantages of having both review types as part of the formal program-evaluation process.

The proliferation of CS programs and the common goal of many CS programs to collect critical long-term monitoring data underscore two opportunities. First, there is an opportunity for the field of CS to play a larger role in informing conservation and resource management. Second, there is a need for a formal CS program-evaluation process to ensure that the data from these programs are usable in these contexts. Structured and periodic program evaluations are needed to identify the strengths and limitations of CS programs and to support the integration of data from these programs into applied conservation and management. This integration will allow CS programs to accomplish their stated goals and objectives, ensure that monitoring resources are used efficiently and effectively, and increase the return on investment from citizen-led monitoring efforts.

Acknowledgments

This manuscript would not have been possible without direct cooperation and engagement by the San Diego Tracking

Team. We also benefited from discussions with the San Diego Citizen Science Network.

Funding

CAT, RLL, and DHD were supported in part by Local Assistance Grant #PO982020 from the San Diego Association of Governments (SANDAG). RLL was supported in part by NASA Earth Science Division/Applied Sciences Program's ROSES-2012 Ecological Forecasting grant (NNH12ZDA001N-COF) and by the Natural Community Conservation Planning Program.

Supplemental material

Supplementary data are available at *BIOSCI* online.

References cited

- Aceves-Bueno E, et al. 2015. Citizen science as an approach for overcoming insufficient monitoring and inadequate stakeholder buy-in in adaptive management: criteria and evidence. *Ecosystems* 18: 493.
- Adamcik RS, Bellantoni ES, DeLong DC Jr., Schomaker JH, Hamilton DB, Laubhan MK, Schroeder RL. 2004. Writing Refuge Management Goals and Objectives: A Handbook. US Fish and Wildlife Service.
- Atkinson PW, et al. 2006. Identifying declines in waterbirds: The effects of missing data, population variability and count period on the interpretation of long-term survey data. *Biological Conservation* 130: 549–559.
- Beaubien EG, Hamann A. 2011. Plant phenology networks of citizen scientists: Recommendations from two decades of experience in Canada. *International Journal of Biometeorology* 55: 833–841.
- Beirne C, Lambin X. 2013. Understanding the determinants of volunteer retention through capture–recapture analysis: Answering social science questions using a wildlife ecology toolkit. *Conservation Letters* 6: 391–401.
- Bell S, Marzano M, Reinert H, Cent J, Kobierska H, Podjed D, Vandzinskaite D, Armaitiene A, Grodzińska-Jurczak M, Muršič R. 2007. The Social Science of Participatory Monitoring Networks. Durham University. EuMon Project no. 006463. (1 January 2015; http://eumon.ckff.si/deliverables_public/D24e.pdf)
- Bell S, Marzano M, Cent J, Kobierska H, Podjed D, Vandzinskaite D, Reinert H, Armaitiene A, Grodzińska-Jurczak M, Muršič R. 2008. What counts? Volunteers and their organizations in the recording and monitoring of biodiversity. *Biodiversity Conservation* 7: 3443–3454.
- Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J. 2009. Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience* 59: 977–984.
- Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK. 2014. Next steps for citizen science. *Science* 343: 1436–1438.
- Bonter DN, Cooper CB. 2012. Data validation in citizen science: A case study from Project FeederWatch. *Frontiers in Ecology and the Environment* 10: 305–307.
- Bonter DN, Harvey MG. 2008. Winter survey data reveal rangewide decline in Evening Grosbeak populations. *Condor* 110: 376–381.
- Brereton T, Roy DB, Middlebrook I, Botham M, Warren M. 2011. The development of butterfly indicators in the United Kingdom and assessments in 2010. *Journal of Insect Conservation* 15: 139–151.
- Burton AC. 2012. Critical evaluation of a long-term, locally based wildlife monitoring program in West Africa. *Biodiversity Conservation* 21: 3079–3094.
- Chandler M, et al. 2016. Contribution of citizen science toward international biodiversity monitoring. *Biological Conservation*. (19 July 2017; <http://dx.doi.org/10.1016/j.biocon.2016.09.004>)
- Chase SK, Levine A. 2016. A framework for evaluating and designing citizen science programs for natural resources monitoring. *Conservation Biology* 30: 456–466.

- Connors JP, Lei SF, Kelly M. 2012. Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Annals of the Association of American Geographers* 102: 1267–1289.
- Conrad CC, Hilchey KG. 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring Assessment* 176: 273–291.
- Conrad CT, Daoust T. 2008. Community-based monitoring frameworks: Increasing the effectiveness of environmental stewardship. *Environmental Management* 41: 358–366.
- Constantino PDAL, Carlos HSA, Ramalho EE, Rostant L, Marinelli CE, Teles D, Fonseca-Junior SF, Fernandes RB, Valsecchi J. 2012. Empowering local people through community-based resource monitoring: A comparison of Brazil and Namibia. *Ecology and Society* 17 (art. 22). (19 July 2017; <http://dx.doi.org/10.5751/ES-05164-170422>)
- Cook CB, Possingham HP, Fuller RA. 2013. Contribution of systematic reviews to management decisions. *Conservation Biology* 27: 902–915.
- Cooper CB, Dickinson J, Phillips T, Bonney R. 2007. Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society* 12 (art. 11).
- Cooper CB, Shirk J, Zuckerman B. 2014. The invisible prevalence of citizen science in global research: Migratory birds and climate change. *PLOS ONE* 9 (art. e106508).
- Couvet D, Jiguet F, Julliard R, Levrel H, Teysseire A. 2008. Enhancing citizen contributions to biodiversity science and public policy. *Interdisciplinary Science Reviews* 33: 95–103.
- Cox TE, Philippoff J, Baumgartner E, Smith CM. 2012. Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecological Applications* 22: 1201–1212.
- Crain R, Cooper C, Dickinson JL. 2014. Citizen science: A tool for integrating studies of human and natural systems. *Annual Review of Environment and Resources* 39: 641–665.
- Crall AW, Newman GJ, Jarnevich CS, Stohlgren TJ, Waller DM, Graham J. 2010. Improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions* 12: 3419–3428.
- Danielsen F, Burgess ND, Balmford A. 2005. Monitoring matters: Examining the potential of locally-based approaches. *Biodiversity and Conservation* 14: 2507–2542.
- Danielsen F, Mendoza MM, Tagtag A, Alviola PA, Balet DS, Jensen AE, Enghoff M, Poulson MK. 2007. Increasing conservation management action by involving local people in natural resource monitoring. *Ambio* 36: 566–570.
- Danielsen F, et al. 2009. Local participation in natural resource monitoring: A characterization of approaches. *Conservation Biology* 23: 31–42.
- Danielsen F, et al. 2014. A multicountry assessment of tropical resource monitoring by local communities. *BioScience* 64: 236–251.
- DeBlust G, Laurijssens G, H. Van Calster H, Verschelde P, Bauwens D, De Vos B, Svensson J. 2012. Design of a Monitoring System and Its Cost Effectiveness: Optimization of Biodiversity Monitoring through Close Collaboration of Users and Data Providers. EBONE, the European Biodiversity Observation Network. Report no. 8.1.
- De Solla SR, Shirose LJ, Fernie KJ, Barrett GC, Brousseau CS, Bishop CA. 2005. Effect of sampling effort and species detectability on volunteer based anuran monitoring programs. *Biological Conservation* 121: 585–594.
- Devictor V, Wittaker RJ, Beltrame C. 2010. Beyond scarcity: Citizen science programmes as useful tools for conservation biogeography. *Diversity and Distributions* 16: 354–362.
- Dickinson JL, Zuckerman B, Bonter DN. 2010. Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics* 41: 149–172.
- Doerr ED, Dorrrough J, Davis MJ, Doerr VAJ, McIntyre S. 2015. Maximizing the value of systematic reviews in ecology when data or resources are limited. *Austral Ecology* 40: 1–11.
- Dolreny S, Hazzah L, Frank LG. 2016. Conservation and monitoring of a persecuted African lion population by Maasai warriors. *Conservation Biology* 30: 467–475.
- Donnelly A, Crowe O, Regan E, Begley S, Caffarra A. 2014. The role of citizen science in monitoring biodiversity in Ireland. *International Journal of Biometeorology* 58: 1237–1249.
- Etienne M, DuToit DR, Pollard S. 2011. ARDI: A co-construction method for participatory modeling in natural resources management. *Ecology and Society* 16 (art. 44).
- Eyre TJ, Fisher A, Hunt LP, Kutt AS. 2011. Measure it to better manage it: A biodiversity monitoring framework for the Australian rangelands. *Rangeland Journal* 33: 239–253.
- Forrester G, Baily P, Conetta D, Forrester L, Kintzing E, Jarecki L. 2015. Comparing monitoring data collected by volunteers and professionals shows that citizen scientists can detect long-term change on coral reefs. *Journal for Nature Conservation* 24: 1–9.
- Framstad E, Henle K, Henry P, Lengyel S, Marzano M, Nowicki P, Schmeller D. 2008. Best practice for monitoring species and habitats of community interests. Helmholtz Center for Environmental Research. EuMon Project no. 006463. (1 January 2015; http://eumon.ckff.si/deliverables_public/D30.pdf)
- Freitag A, Pfeffer MJ. 2013. Process, not product: Investigating recommendations for improving Citizen Science “success.” *PLOS ONE* 8 (art. e64079).
- Funder M, Danielsen F, Ngaga Y, Nielsen MR, Poulson MK. 2013. Reshaping conservation: The social dynamics of participatory monitoring in Tanzania’s community-managed forests. *Conservation and Society* 11: 218–232.
- Gallo T, Waitt D. 2011. Creating a successful citizen science model to detect and report invasive species. *BioScience* 61: 459–465.
- Gollan J, de Bruyn LL, Reid N, Wilkie L. 2012. Can volunteers collect data that are comparable to professional scientists? A study of variables used in monitoring the outcomes of ecosystem rehabilitation. *Environmental Management* 50: 969–978.
- Gouveia C, Fonseca A, Camara A, Ferreira F. 2004. Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management* 71: 135–154.
- Greenwood JJD. 2007. Citizens, science and bird conservation. *Journal of Ornithology* 148 (suppl. 1): 77–124.
- Havens K, Henderson S. 2013. Citizen science takes root: Building on a long tradition, amateur naturalists are gathering data for understanding both seasonal events and the effects of climate change. *American Scientist* 101: 378.
- Hochachka WM, Fink D, Hutchinson RA, Sheldon D, Wong WK, Kelling S. 2012. Data-intensive science applied to broad-scale citizen science. *Trends in Ecology and Evolution* 27: 130–137.
- Hunter J, Alabri A, van Ingen C. 2013. Assessing the quality and trustworthiness of citizen science data. *Concurrency and Computation: Practice and Experience* 25: 454–466.
- Jiguet F, Devictor V, Julliard R, Couvet D. 2012. French citizens monitoring ordinary birds provides tools for conservation and ecological sciences. *Acta Oecologica* 44: 58–66.
- Jordan RC, Gray SA, Howe DV, Brooks WR, Ehrenfeld JG. 2011. Knowledge gain and behavioral change in citizen-science programs. *Conservation Biology* 25: 1148–1154.
- Jordan RC, Ballard HL, Phillips TB. 2012a. Key issues and new approaches for evaluating citizen-science learning outcomes. *Frontiers in Ecology and the Environment* 10: 307–309.
- Jordan RC, Brooks WR, Howe DV, Ehrenfeld JG. 2012b. Evaluating the performance of volunteers in mapping invasive plants in public conservation lands. *Environmental Management* 49: 425–434.
- Jordan R, Crall A, Gray S, Phillips T, Mellor D. 2015. Citizen science as a distinct field of inquiry. *BioScience* 65: 208–211.
- Kremen C, Ullmann KS, Thorp RW. 2011. Evaluating the quality of citizen-scientist data on pollinator communities. *Conservation Biology* 25: 607–617.
- Lindenmayer DB, Likens GE. 2010. The science and application of ecological monitoring. *Biological Conservation* 143: 1317–1328.
- Lorkie CJ. 2014. Formalized synthesis opportunities for ecology: Systematic reviews and meta-analyses. *Oikos* 123: 897–902.

- Lukyanenko R, Parsons J, Wiersma YF. 2016. Emerging problems of data quality in citizen science. *Conservation Biology* 30: 447–449.
- Mackechnie C, Maskell L, Norton L, Roy D. 2011. The role of “big society” in monitoring the state of the natural environment. *Journal of Environmental Monitoring* 13: 2687–2691.
- Margoluis R, Stem C, Salafsky N, Brown M. 2009. Using conceptual models as a planning and evaluation tool in conservation. *Evaluation and Program Planning* 32: 138–147.
- Marshall NJ, Kleine DA, Dean AJ. 2012. CoralWatch: Education, monitoring, and sustainability through citizen science. *Frontiers in Ecology and the Environment* 10: 332–334.
- Matteson KC, Taron DJ, Minor ES. 2012. Assessing citizen contributions to butterfly monitoring in two large cities. *Conservation Biology* 26: 557–564.
- McKinley DC, et al. 2016. Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation* 208: 15–28.
- Moyer-Horner L, Smith MM, Belt J. 2012. Citizen science and observer variability during American pika surveys. *Journal of Wildlife Management* 76: 1472–1479.
- Mullen MW, Allison BE. 1999. Stakeholder involvement and social capital: Keys to watershed management success in Alabama. *Journal of the American Water Resources Association* 35: 655–662.
- Nerbonne JF, Nelson KC. 2008. Volunteer macroinvertebrate monitoring: Tensions among group goals, data quality, and outcomes. *Environmental Management* 42: 470–479.
- Newman G, Graham J, Crall A, Laituri M. 2011. The art and science of multi-scale citizen science support. *Ecological Informatics* 6: 217–227.
- Newman G, Wiggins A, Crall A, Graham E, Newman S, Crowston K. 2012. The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment* 10: 298–304.
- Phillips TB, Ferguson M, Minarchek M, Porticella N, Bonney R. 2014. User’s Guide for Evaluating Learning Outcomes in Citizen Science. Cornell Lab of Ornithology.
- Pocock MJO, Chapman DS, Sheppard LJ, Roy HE. 2014. A Strategic Framework to Support the Implementation of Citizen Science for Environmental Monitoring: Final Report to SEPA. Centre for Ecology and Hydrology.
- Possingham HP, Wintle BA, Fuller RA, Joseph LN. 2012. The conservation return on investment from ecological monitoring. Pages 49–61 in Lindenmayer DB, Gibbons P, eds. *Biodiversity Monitoring in Australia*. CSIRO.
- Powell MC, Colin M. 2008. Meaningful citizen engagement in science and technology: What would it really take? *Science Communication* 30: 126–136.
- Predavec M, Lunney D, Hope B, Stalenberg E, Shannon I, Crowther MS, Miller I. 2016. The contribution of community wisdom to conservation ecology. *Conservation Biology* 30: 496–505.
- Pullin AS, Stewart GB. 2006. Guidelines for systematic review in conservation and environmental management. *Conservation Biology* 20: 1647–1656.
- Reed MS. 2008. Stakeholder participation for environmental management: A literature review. *Biological Conservation* 141: 2417–2431.
- Riesch H, Potter C. 2014. Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science* 23: 107–120.
- Schmeller DS, et al. 2008. Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology* 23: 307–316.
- Schmiedel U, et al. 2016. Contributions of paraecologists and parataxonomists to research, conservation, and social development. *Conservation Biology* 20: 506–519.
- Schroeder RL. 2009. Evaluating the quality of biological objectives for conservation planning in the National Wildlife Refuge System. *George Wright Forum* 26: 22–30.
- Sergeant CJ, Moynahan BJ, Johnson WJ. 2012. Practical advice for implementing long-term ecosystem monitoring. *Journal of Applied Ecology* 49: 969–973.
- Sharpe A, Conrad C. 2006. Community based ecological monitoring in Nova Scotia: Challenges and opportunities. *Environmental Monitoring and Assessment* 113: 395–409.
- Shirk J, Bonney R. 2015. Citizen Science Framework Review: Informing a Framework for Citizen Science within the US Fish and Wildlife Service. Cornell Lab of Ornithology.
- Shirk JL, et al. 2012. Public participation in scientific research: A framework for deliberate design. *Ecology and Society* 17 (art. 29).
- Silvertown J. 2009. A new dawn for citizen science. *Trends in Ecology and Evolution* 24: 467–471.
- Sultana P, Abeyasekera S. 2008. Effectiveness of participatory planning for community management of fisheries in Bangladesh. *Journal of Environmental Management* 86: 201–213.
- Tear TH, et al. 2005. How much is enough? The recurrent problem of setting measurable objectives in conservation. *BioScience* 55: 835–849.
- Theobald EJ, et al. 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation* 181: 236–244.
- Tulloch AIT, Szabo JK. 2012. A behavioural ecology approach to understand volunteer surveying for citizen science datasets. *Emu: Austral Ornithology* 112: 313–325.
- Tulloch AIT, Possingham HP, Joseph LN, Szabo J, Martin TG. 2013. Realising the full potential of citizen science monitoring programs. *Biological Conservation* 165: 128–138.
- Van der Wal R, Sharma N, Mellish C, Robinson A, Siddharthan A. 2016. The role of automated feedback in training and retaining biological recorders for citizen science. *Conservation Biology* 30: 550–561.
- Westgate MJ, Likens GE, Lindenmayer DB. 2013. Adaptive management of biological systems: A review. *Biological Conservation* 158: 128–139.
- Whitelaw G, Vaughan H, Craig B, Atkinson D. 2003. Establishing the Canadian community monitoring network. *Environmental Monitoring and Assessment* 88: 409–418.
- Yoccoz NG, Nichols JD, Boulinier T. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16: 446–453.

Catherine A. Tredick is affiliated with Stockton University, in Galloway, New Jersey. Rebecca L. Lewison and Douglas H. Deutschman are affiliated with San Diego State University, in California. Timothy Ann Hunt, Karen L. Gordon, and Phoenix Von Henty are with the San Diego Tracking Team, in California.