LETTER

Assessing citizen science data quality: an invasive species case study

Alycia W. Crall[1], Gregory J. Newman[1], Thomas J. Stohlgren[2], Kirstin A. Holfelder[1], Jim Graham[1], & Donald M. Waller[3]

[1] Natural Resource Ecology Laboratory, Colorado State University, Campus Delivery 1499, Fort Collins, CO 80523, USA
[2] Fort Collins Science Center, US Geological Survey, 2150 Centre Avenue Bldg C, Fort Collins, CO 80526, USA
[3] Department of Botany, University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706, USA

Keywords
Citizen science; data quality; invasive species; non-native species; vegetation monitoring; volunteer monitoring protocols.

Correspondence
Alycia W. Crall, Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523–1499, USA. Tel: +(970) 491-2302; Fax: +(970) 491-1965. E-mail: Alycia.Crall@colostate.edu

Received
11 April 2011
Accepted
5 July 2011

Editor
Edward Webb

doi: 10.1111/j.1755-263X.2011.00196.x

Abstract

An increase in the number of citizen science programs has prompted an examination of their ability to provide data of sufficient quality. We tested the ability of volunteers relative to professionals in identifying invasive plant species, mapping their distributions, and estimating their abundance within plots. We generally found that volunteers perform almost as well as professionals in some areas, but that we should be cautious about data quality in both groups. We analyzed predictors of volunteer success (age, education, experience, science literacy, attitudes) in training-related skills, but these proved to be poor predictors of performance and could not be used as effective eligibility criteria. However, volunteer success with species identification increased with their self-identified comfort level. Based on our case study results, we offer lessons learned and their application to other programs and provide recommendations for future research in this area.

Introduction

Citizen science represents a partnership between volunteers and scientists to address research questions. These partnerships have expanded in number and scope as a way to connect scientific research to public outreach and education while providing additional resources to professional surveys (Bonney *et al.* 2009; Lepczyk *et al.* 2009). Data collected by citizen scientists inform natural resource management (Brown *et al.* 2001), environmental regulation (Penrose and Call 1995), and scientific research (Cooper *et al.* 2007). Therefore, data quality is paramount and could have far-reaching environmental, social, and/or political implications (Engel and Voshell 2002).

Several studies have examined data quality in citizen science programs by determining predictors of participant success (Danielsen *et al.* 2005). Accuracy rates within these programs tend to vary, and results are rarely made available to the larger citizen science community. Standardizing monitoring protocols, designed by professionals and field-tested with citizen scientists working under realistic conditions, can improve data quality and analyses (Delaney *et al.* 2008).

We tested the ability of volunteers to conduct an invasive plant species monitoring protocol following 1 day of training. We tested participants' ability to identify species and implement the protocol compared with professionals to determine eligibility criteria by examining which factors were most strongly associated with performance. To our knowledge, no other study has yet used social predictors to assess success in such programs.

Methods

Participant recruitment

We recruited participants from existing volunteer networks (typical of citizen science programs) and provided them with a short survey. We recorded demographic information and willingness to attend the training or serve in a control group. This resulted in self-selection of participants, so we also collected demographic information from nonparticipants ($N = 166$) to ensure that our experimental group adequately represented the volunteer population.

In 2009, we held trainings at the University of Wisconsin Arboretum, Madison, Wisconsin ($N = 31$), and at Colorado State University's Environmental Learning Center, Fort Collins, Colorado ($N = 28$). University professors, graduate students, and land managers working with invasive plants participated as professionals (WI: $N = 31$; CO: $N = 21$). Some professionals were not plant taxonomists, but all had extensive field experience with invasive plant species prior to participation.

Evaluation

We administered evaluations to participants using the pretest/posttest control group design, with the control group (untrained participants) receiving a pretraining evaluation only (Campbell and Stanley 1963). The pretraining evaluation collected information on participant demographics, personal behavior and engagement, science literacy, and attitudes toward the environment and technology (Friedman 2008). Participants rated their level of experience with certain skills: no experience, little experience, some experience, proficient, or expert (Table 1).Personal behavior and engagement were assessed using statements related to how frequently individuals participate in various activities: never, a few times each year, each month, every week, or every day (Table 1). Participants also noted their comfort in identifying the plant species taught during the training using a five-point response: very uncomfortable to very comfortable.

We adopted standard scales to assess science literacy, attitude toward the environment, and attitude toward technology (Brossard *et al.* 2005). Participants responded to the standard science and engineering indicators question: "Please tell us in your own words what it means to study something scientifically" (National Science Board 1996). We asked two additional questions related to science literacy but specific to the content covered in our training: "write a research question that can be answered by collecting data on invasive species" and "how would you set up a sampling design to answer this research question?"

Attitude toward the environment was assessed with a subset of the new environmental paradigm scale (frequently used measure of public environmental concern; Stern *et al.* 1995) with the addition of several additional scale items selected following the guidelines of Marcinkowski (1997; Table 1). Responses ranged from strongly disagree to strongly agree. We assessed attitude toward technology with the computer attitude scale using the same five responses (Table 1; Nickell and Pinto 1986).

Training workshops

The training included introductions to invasive species, plant identification, GPS use, sampling design, and a sampling protocol (see section "Sampling Protocol for Invasive Plant Species"). We selected six species at each training site for species identification training. Species selected included three easily recognizable species and three that could be easily confused with other species (Table 2). Although subjective, these classifications were based on experience with how easily volunteers can identify these species in the field.

Sampling protocol for invasive plant species

Barnett *et al.* (2007) described a sampling design for invasive plant species that integrates mapping techniques with a 168-m^2 plot to monitor species distributions and abundance (Fig. 1). The protocol generates data comparable to the U.S. Forest Service's Forest Inventory and Analysis Program, the National Institute of Invasive Species Science, the U.S. Fish and Wildlife Service's Invasive and Volunteers Program, and the National Ecological Observatory Network. We modified this protocol to accommodate volunteers. To map species, we asked participants to record point locations only. For plot-based assessments, participants recorded data for those species they had been taught to identify. Within each plot (Fig. 1), volunteers recorded presence of each target species and cover of each shrub and tree species. S(he) also estimated herbaceous cover in three vegetation of 1-m^2 subplots (Fig. 1).

Skills testing

Volunteers performed four tasks (plant identification, GPS use, plot implementation, plot setup) to test skills taught in the training. The first tested volunteers' ability to identify the six target plant species. We tagged 125 plants (target and nontarget species) along established trails with an identification number. Volunteers and professionals walked through each search area, identifying target species as they were encountered. Volunteers also recorded waypoints (i.e., datum, zone, universal transverse mercator [UTM] easting, UTM northing, accuracy) for five marked stakes. To test navigation skills, participants recorded the name of the stake marking a waypoint saved in their GPS unit. Five plots were set up

Table 1 Results of reliability analyses (item total correlation, alpha if item deleted, Cronbach's alpha) for experience, attitude toward the environment, and attitude toward technology indices. Indices were computed by summing the response for each item after reversed items had been recoded. Statements included in the new environmental paradigm scale are marked with an asterisk

Index	Item total correlation	Alpha if item deleted	Cronbach's alpha
Experience index: skills, personal behavior, and engagement			0.88
Vegetation sampling design	0.56	0.88	
Plant identification	0.67	0.87	
Invasive plant identification	0.77	0.86	
Vegetation monitoring	0.68	0.87	
Volunteering for environmental organizations	0.44	0.89	
Attending community events related to environmental issues	0.52	0.88	
Removing/controlling invasive species	0.70	0.87	
Monitoring invasive species	0.69	0.87	
Educating others about invasive species	0.73	0.86	
Attitude Index: Environment			0.82
*Humans were created to rule over the rest of nature.	0.67	0.79	
*People have the right to modify the natural environment to suit their needs.	0.44	0.82	
*Plants and animals exist primarily to be used by people.	0.69	0.79	
*People need not to adapt to the natural environment because they can remake it to suit their needs.	0.49	0.81	
I think most of the concern about environmental problems and issues is important to me.	0.42	0.82	
I am concerned about the issue of invasive species.	0.29	0.83	
There are already enough laws to protect the environment.	0.57	0.80	
I would oppose any environmental regulations that would restrict my way of life.	0.59	0.80	
Attitude index: technology			0.81
People are becoming slaves to computers.	0.49	0.79	
Computers make me uncomfortable because I don't understand them.	0.42	0.79	
The use of computers is enhancing our standard of living.	0.56	0.78	
Computers are responsible for many of the good things we enjoy.	0.44	0.79	
Soon our lives will be controlled by computers.	0.40	0.80	
Computers can eliminate a lot of tedious work for people.	0.32	0.80	
There are unlimited possibilities of computer applications that have not even been thought of yet.	0.43	0.79	
I feel intimidated by computers.	0.53	0.78	
Computers are lessening the importance of too many jobs now done by humans.	0.54	0.78	
Computers are a fast and efficient means of gaining information.	0.47	0.79	
Life will be easier and faster with computers.	0.40	0.80	
Computers are difficult to understand and frustrating to work with.	0.40	0.80	

at each training site to test implementation of the plot protocol. For each plot and its three nested subplots, volunteers and professionals recorded presence/absence and abundance data (Fig. 1). An evaluator recorded volunteers' ability to set up the plot correctly in the field.

Statistical analyses

Using SPSS (Version 18; 2009), we transformed data, where needed for normative residuals. We calculated chi-square statistics to compare demographics between control and treatment groups. We compared the abilities of volunteers and professionals to correctly identify species and to identify species by difficulty using chi-square analyses. We used regression to assess correlations between (1) volunteer accuracy and professional accuracy in species identification and (2) volunteer accuracy and volunteer comfort level identifying species. *Rhamnus cathartica* and *R. frangula* were merged for this and similar analyses because comfort level was identified only to the genus.

We scored GPS skills for each participant using points for correctly identifying the datum and UTM zone and navigating to the appropriate marker. Volunteers received points if waypoints were located within a 15-m buffer zone of the marker. Because each participant recorded five locations, these scores were divided by two to weight their importance relative to other GPS skills. GPS scores ranged between zero and 5.5.

We also compared the ability of volunteers and professionals to correctly note the presence of "easy" and "difficult" species in each plot. We used regression to

Table 2 Six species taught during the two trainings in Wisconsin and Colorado, including identification difficulty classification

Scientific name	Common name	State	Identification difficulty
R. cathartica L.	Common Buckthorn	WI	Easy
Hesperis matronalis L.	Dame's Rocket	WI	Easy
Alliaria petiolata (M. Bieb.) Cavara & Grande	Garlic Mustard	WI	Easy
R. frangula Mill.	Glossy Buckthorn	WI	Difficult
C. orbiculatus Thunb.	Asian Bittersweet	WI	Difficult
Lonicera sp. L.	Honeysuckle	WI	Difficult
E. esula L.	Leafy Spurge	CO	Easy
Linaria dalmatica (L.) Mill.	Dalmation Toadflax	CO	Easy
Elaeagnus angustifolia L.	Russian Olive	CO	Easy
Carduus nutans L.	Musk Thistle	CO	Difficult
Cynoglossum officinale L.	Houndstongue	CO	Difficult
Cardaria draba (L.) Desv.	Whitetop	CO	Difficult

assess the ability of volunteers to correctly record species presence compared to professionals. We also calculated correlations between volunteers' comfort level identifying species and their accuracy at recording presence. We ran a two-way ANOVA to look at significant differences in cover estimates attributable to group, species, and the interaction between these two factors.

We tested several social variables for their ability to predict participant success at each skill. Each participant's science literacy score was based on responses to three science literacy questions. These were scored as valid or invalid by three coders (Lacy and Riffe 1996). This process produced an adequate interrater reliability for science literacy responses ($\alpha = 0.69, 0.77, 0.70$; Krippendorff 2004). Total science literacy scores ranged from zero to five.

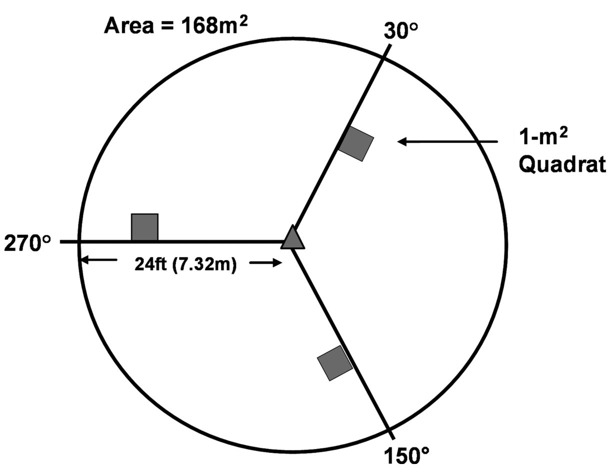

Figure 1 The modified Forest Inventory Analysis (FIA) plot (see Barnett *et al.* 2007) served as part of our invasive plant species monitoring protocol for citizen scientists.

We generated experience, attitude toward the environment, and attitude toward technology indices from responses to several Likert scale questions from the pretraining survey (Table 1). We tested for significant differences in scores and indices between treatment and control volunteers using t-tests. We used regression tree analyses to evaluate which predictors (age, education, science literacy score, experience index, attitude towards the environment index) best predicted success at all but one skill. We used multiple regression based on age, education, GPS experience, and attitude towards technology to predict trainees' GPS score. As these variables had low correlations among themselves, all were included in analyses.

Results

Participant demographics

Our demographic results (Table 3) confirm that data can be combined across states for both treatment ($N = 59$) and control ($N = 110$) groups and that participants represented the sample population fairly well. Once merged, we found no significant difference between treatment and control groups for science literacy ($t = 0.07; p = 0.95$), experience ($t = -1.4$; $p = 0.15$), attitude toward the environment ($t = -0.66$; $p = 0.51$), or attitude toward technology ($t = -1.3$; $p = 0.21$).

Performance with monitoring skills

Professionals identified species more accurately than volunteers (88% vs. 72%; $\chi^2 = 104$; $p < 0.01$) who had 28% false negative (species not identified when present) and 1% false positive (species identified when not present) identifications, respectively (vs. 12% and <1% by the professionals; Table 4). Volunteers correctly identified "easy" species 82% of the time versus 65% for "difficult" species ($\chi^2 = 0.19$; $p = 0.67$; Table 4). Self-identified comfort levels with identifying species proved fairly accurate in predicting correct identification ($r = 0.69$; $p = 0.04$). Across species, volunteer accuracy strongly paralleled the accuracy of professionals in identifying species ($r = 0.96$; $p < 0.01$).

Although 69% of volunteers had little or no prior experience using GPS, their average success after training was 75%. Most (85%) volunteers recorded the UTM zone correctly and 84% recorded marker coordinates correctly. Recording the datum and navigating proved more difficult with 64% and 67% correct, respectively (mean GPS score: 4.2 ± 0.17).

Professionals were more accurate in recording the presence/absence of species within plots than volunteers

Table 3 Demographics of participants (control and treatment groups) and nonparticipants. No significant differences existed between control and treatment groups at each study site and nonparticipants in age, gender, education, or profession. Income did differ ($\chi^2 = 16$; $p < 0.01$) with participants having higher income households. In Wisconsin, no differences existed between control and treatment groups for any demographic variable, but in Colorado, the control group was younger than the treatment group ($\chi^2 = 11$; $p = 0.05$). Across both states, the control group had a greater percentage of higher income households compared to the treatment group ($\chi^2 = 7.8$; $p = 0.05$). Differences in gender were almost significant ($\chi^2 = 3.6$; $p = 0.06$) with a greater percentage of females in the treatment than control group. For those categories that the percentages do not add up to 100%, the remaining portion did not respond. NA indicates no data available

		Colorado		Wisconsin	
	Nonparticipants	Control	Treatment	Control	Treatment
Age					
18–24	15	3	4	7	6
25–34	13	26	0	11	3
35–44	8	18	14	10	16
45–54	10	20	18	16	16
55–64	27	24	50	32	29
65–75	21	9	14	22	29
Over 75	6	0	0	3	0
Gender					
Male	42	35	18	41	29
Female	58	65	79	59	68
Education					
Less than high school	0	0	0	0	0
High school/GED	0	3	0	3	3
Some college	17	6	14	7	16
2-Year college degree (associates)	4	9	0	3	3
4-Year college degree (BA, BS)	38	47	50	41	35
Master's degree	33	23	25	30	32
Doctoral degree	6	9	11	12	6
Professional degree (MD, JD)	2	3	0	4	3
Profession					
Student (nonscience emphasis)	NA	0	0	1	0
Student (science emphasis)	NA	9	0	3	13
Professor/teacher (nonscience emphasis)	NA	9	7	16	6
Professor/teacher (science emphasis)	NA	9	10	11	16
Scientist/engineer	NA	9	10	12	3
Land manager	NA	15	7	4	6
IT professional	NA	8	4	3	3
Volunteer coordinator	NA	0	4	4	6
Nonprofit administrator	NA	0	4	0	3
Nonprofit employee	NA	3	0	10	0
Other	NA	38	54	36	42
Income					
1 person working	25	26	43	32	29
2 or more people working	33	56	43	42	39
Savings/investments	21	9	7	21	23
Other	17	0	7	0	6

(91% vs. 82%; $\chi^2 = 50$; $p < 0.01$; Table 5). Species identification difficulty affected accuracy ($\chi^2 = 54$; $p < 0.01$), and volunteer and professional accuracy were correlated ($r = +0.68$; $p = 0.02$). Comfort level with species identification failed to predict volunteer success in recording species' presence/absence ($R^2 = 0.01$; $p = 0.80$).

Estimates of cover across species within plots did not differ between volunteers and professionals ($t = 1.8$; $p = 0.08$). Estimates of percent cover depended on species ($F = 172$; $p < 0.01$) and the interaction between species and group ($F = 3.9$; $p < 0.01$). Estimates of subplot cover also depended on species ($F = 15$; $p < 0.01$), but was not dependent on group ($F = 0.34$; $p = 0.56$) or the

Table 4 The frequency of species within the search areas (out of 125 marked species) and the volunteer group's average comfort level with these species prior to the training (standard error in parentheses). Percent correct and false positive and false negative identifications are also shown for the volunteers and professionals. "Easy" species are marked with an asterisk. The average across species, shown in bold type, includes both target and nontarget species

				Volunteers			Professionals		
Species	State	Frequency of Species	Average comfort level	Correct (%)	False positive (%)	False negative (%)	Correct (%)	False positive (%)	False negative (%)
*R. cathartica**	WI	23	3.3 (0.04)[a]	74	2	26	87	1	11
*H. matronalis**	WI	0	3.2 (0.04)	NA	<1	NA	NA	<1	NA
*A. petiolata**	WI	16	4.0 (0.03)	83	1	17	90	<1	10
C. orbiculatus	WI	5	2.0 (0.03)	72	3	28	91	1	9
R. frangula	WI	1	3.3 (0.04)[a]	27	2	73	63	3	37
Lonicera sp.	WI	15	3.4 (0.04)	80	1	20	89	<1	11
*L. dalmatica**	CO	4	2.8 (0.04)	82	1	18	95	<1	5
*E. esula**	CO	26	3.0 (0.04)	90	2	10	96	1	4
*E. angustifolia**	CO	0	4.5 (0.02)	NA	<1	NA	NA	0	NA
C. officinale	CO	1	2.2 (0.04)	67	1	33	90	<1	10
C. nutans	CO	21	2.9 (0.04)	87	1	13	94	<1	6
C. draba	CO	1	1.8 (0.04)	54	2	46	80	<1	20
Average Across Species				**72**	**1**	**28**	**88**	**<1**	**12**

[a]Comfort level for *Rhamnus* species was identified to the genus level only.

interaction between species and group ($F = 0.72$; $p = 0.49$; Fig. 2). Only 60% of the participants set the plot up correctly in the field with many errors related to compass use (70%).

Predictors of performance

Regression analyses revealed no significant predictors for success at identifying species, observing species presence, or success at GPS skills. Participants younger than 45 performed better at plot setup, but the regression tree proportional education in error value (similar to an R^2 value in regression) was low (0.15).

Discussion

Citizen science has the capacity to provide additional resources for professional monitoring activities, improve collaboration, and promote education (Bonney *et al.* 2009), allowing such programs to contribute significantly

Table 5 The frequency of species found within the five plots and the volunteers' average comfort level with these species prior to the training (standard error in parentheses). Percent correct for recorded presence/absence are also shown for the volunteers and professionals. "Easy" species are marked with an asterisk

Species	State	Frequency of species	Average comfort level	Volunteers correct (%)	Professionals correct (%)
*R. cathartica**	WI	5	3.3 (0.04)[a]	90	98
*H. matronalis**	WI	0	3.2 (0.04)	94	91
*A. petiolata**	WI	4	4.0 (0.03)	72	80
C. orbiculatus	WI	5	2.0 (0.03)	59	96
R. frangula	WI	4	3.3 (0.04)[a]	44	77
Lonicera sp.	WI	4	3.4 (0.04)	75	91
*L. dalmatica**	CO	0	2.8 (0.04)	96	95
*E. esula**	CO	5	3.0 (0.04)	93	93
*E. angustifolia**	CO	1	4.5 (0.02)	98	96
C. officinale	CO	0	2.2 (0.04)	95	95
C. nutans	CO	3	2.9 (0.04)	88	90
C. draba	CO	0	1.8 (0.04)	95	96

[a]Comfort level for *Rhamnus* species was identified to the genus level only.

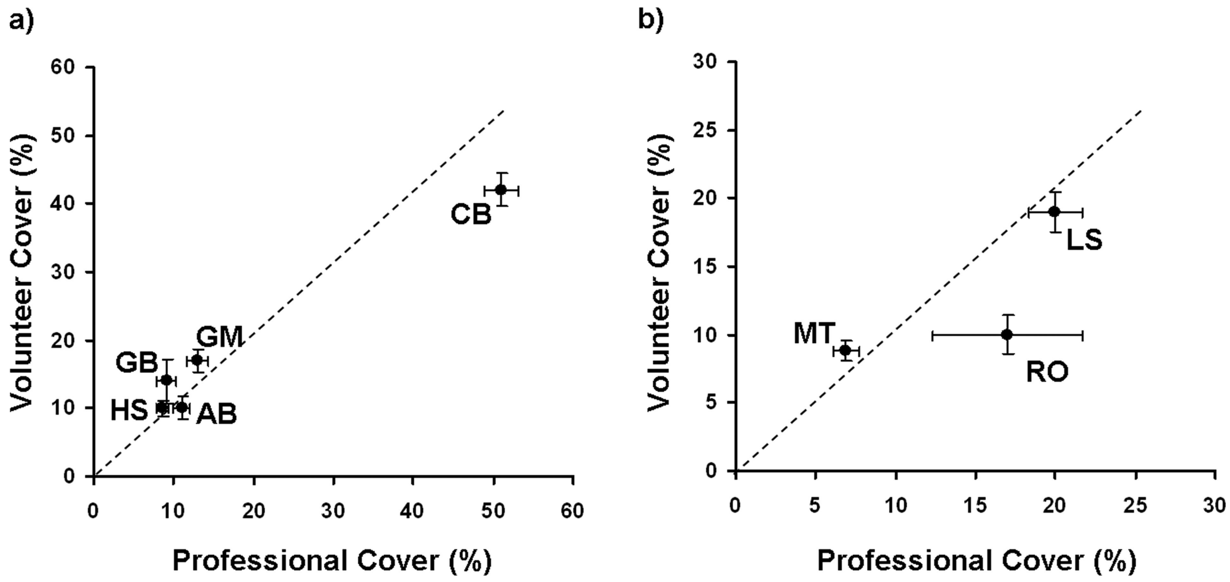

Figure 2 Comparison of cover estimates for all species in Wisconsin (A) and Colorado (B) for volunteers and professionals. Horizontal bars represent standard errors for professionals, and vertical bars represent standard errors for volunteers. Species codes are AB = *C. orbiculatus* (Asian Bittersweet); CB = *R. cathartica* (Common Buckthorn); GM = *A. petiolata* (Garlic Mustard); GB = *R. frangula* (Glossy Buckthorn); HS = *Lonicera* sp. (Honeysuckle); LS = *E. esula* (Leafy Spurge); MT = *C. nutans* (Musk Thistle); and RO = *E. angustifolia* (Russian Olive). "Easy" species are shown in bold type. Only eight species are displayed in the figure because cover estimates were only made on herbaceous species found in the 1-m^2 subplots.

to conservation biology when appropriate protocols are developed and applied across programs. We discuss our results given the limitations of examining this sample after 1 day of training. Our experience index tried to assess learning gains over time based on prior experience/training, but its inability to predict performance suggests additional research is needed in this area. Our study is also limited to the species and conditions specific to our study sites, but these findings can be applied to other programs utilizing volunteer data collectors as described below.

Lessons learned and their applications

Data quality assessments are needed for existing monitoring programs

Our findings underscore the need to test accuracy rates within existing monitoring protocols. The level of accuracy needed will likely depend on the research question being examined and the ability to perform post-hoc statistical manipulation on these data. Although skills testing can be costly and logistically difficult, the results from this and other studies suggest that data quality assumptions cannot be made. Once proper protocols are established, they should be standardized and data quality maintained via regular monitoring of performance to ensure that training and sampling design remain adequate (Danielsen *et al.* 2005).

Species identification training needs to be extended and vouchers included in the protocol

Accurate taxonomic identification requires years of specialized training and remains a barrier of data quality among diverse data collectors. In our study, many volunteers brought skills with them to the training that ranged from a general interest in botany to formal taxonomic training. Although experience has predicted success in some programs (McLaren and Cadman 1999; Nerbonne *et al.* 2008), this was not the case in ours and other studies (Genet and Sargent 2003; Mumby *et al.* 1995). These differences could be attributed to the level of skill required to implement the protocol being tested, the range of experience levels within the study population, or the way in which experience was defined within the study.

Rates of misidentification generally depend on the identification difficulty of a given species. Bloniarz and Ryan (1996) found that volunteers performed better when identifying higher taxonomic categories that differed more dramatically in physical characteristics. In another study, volunteers and professionals differed in recording the frequencies of particular *Ulmus* and *Quercus* species but not when pooled by genus (Brandon *et al.* 2003).

Errors in species identification among data collectors could be handled in numerous ways. Protocols that utilize volunteers could include species for which correct identification rates are high while leaving difficult species to taxonomists. Requiring voucher specimens to be collected for verification could prevent errors but would require expert time to process. Photographing geolocated specimens might be more feasible for checking the accuracy of species identifications, especially if instructions were provided on what physical characteristics need to be photographed to allow efficient taxonomic screening. Smartphone applications, like Leafsnap that uses visual recognition software to help identify tree species from photographs, show promise for advancing accurate species identification.

Volunteers can easily acquire skills to geolocate species

The use of geographic information systems and GPS in natural resource management is growing. Although citizen science programs have begun taking advantage of these technologies (Crall *et al.* 2010), we are unaware of other studies evaluating the accuracy of GPS use by volunteers (but see Jones *et al.* 2008; Rist *et al.* 2010). Most of our participants had little or no experience with GPS use prior to our training but were generally able to use this tool successfully. The main limitation of these technologies may be making them available to large numbers of volunteers (Crall *et al.* 2010). The creation of a network of ″technology libraries″ that host GPS units and other monitoring resources to be checked out for use during volunteer surveys could remedy this.

Volunteers can implement plot-based assessments

Numerous inventories use multiscale plots to assess changes in species distributions and abundance over time (Stohlgren 2007). A network of plot locations randomly placed across a stratified landscape would provide data to more effectively manage plant species. Data from this study indicate that citizen scientists could successfully contribute to such a monitoring network as volunteers and professionals performed similarly in recording presence/absence for most species in field plots.

We often use visual estimates of cover to infer the abundance of plant species, and these estimates usually vary among observers (Kennedy and Addison 1987; Stohlgren 2007; Tonterri 1990). We found that estimated cover values did not differ between groups, and volunteers and professionals tended to under- or overestimate cover differently for different species. Considering the known variability in ocular cover estimates and these findings, we feel volunteers could provide fairly reliable cover estimates if given proper training and a standardized protocol. Similarly, the ability to reliably set up plots should improve with additional compass training.

Recommendations for future research

Assess working in groups

Although we sought to test volunteers independently, another study that grouped volunteers by differing levels of experience proved more successful than our study (Bloniarz and Ryan 1996). Percentage of agreement in tree identification between arborists and volunteer teams (with one member experienced in tree identification) averaged 94% when identified to genus and 80% when identified to species (Bloniarz and Ryan 1996). Grouping volunteers with professionals through established monitoring networks could provide efficient and successful ways to implement monitoring programs (while perhaps adding to its social dimension to make such efforts more attractive).

Assess volunteer certification

Certifying volunteers in particular skills may also prove beneficial. Certification not only provides a feeling of pride and accomplishment for volunteers but also demonstrates his/her long-term commitment to a program, improving volunteer retention (Bell *et al.* 2008). Master naturalists programs established throughout the United States provide a great example of a successful certification model (Main 2004).

Assess technology's role in data quality

Technological advancements can improve the ways we implement successful sampling designs. Several citizen science programs have developed online data entry forms with automated error checking capabilities (Bonney *et al.* 2009; Crall *et al.* 2010). These forms flag suspect data to allow further expert investigation prior to their integration in a widely used dataset. Similarly, smartphone applications have been developed that allow automated entry of location coordinates associated with a species sighting (Crall *et al.* 2010). These sightings could include a photo voucher or an identification tag for a specimen voucher. Although the ability of these tools to improve data quality has not been thoroughly tested, it is likely that they could improve data quality among all data collectors (Stevenson *et al.* 2003; Williams *et al.* 2006).

Determine eligibility criteria for specific skills that can be adopted across programs

Although this study failed to identify reliable predictors of performance among our volunteers, different variables or modifications to our indices may have better predicted volunteer performance. We generated our experience index using a five-point Likert scale. More detailed qualitative data on years of experience with each skill may improve our ability to predict performance as may measures of self-efficacy.

Self-efficacy, defined as an individual's perceived capabilities to perform a specific task (Bandura 1994), has been a cornerstone of psychology research. Its ability to predict performance in work and education settings has been thoroughly reviewed (Judge *et al.* 2007; Usher and Pajares 2008), but its potential role in volunteer data quality needs further examination based on the results found in this study. Our results also underscore the need for more interdisciplinary approaches when addressing these research questions. This information is easy to obtain from volunteers and could be used to efficiently focus training on particular volunteers and species (Genet and Sargent 2003; McLaren and Cadman 1999).

We further recommend that conservation organizations collaborate with scientists developing and evaluating citizen science monitoring networks. Taking advantage of existing assessments helps strengthen existing programs and further standardizes the monitoring protocols already in existence (Bonney *et al.* 2009). As citizen science continues to grow, collaboration among existing and future partners will be essential to its success as a conservation and outreach tool connecting scientists more effectively with the public.

Acknowledgments

This work was funded by the National Science Foundation under grant number OCI-0636213. DMW was supported by the U.S. Department of Agriculture National Research Initiative (CREES Weedy and Invasive Species program, grant no. 2008–35320-18680). Logistical support was provided by the Natural Resource Ecology Laboratory at Colorado State University, the USGS Fort Collins Science Center, the University of Wisconsin-Madison Arboretum, and the Colorado State University Environmental Learning Center. Marian Farrior, Brad Herrick, and Brendon Panke helped coordinate the event at the Arboretum. Catherine Jarnevich, Alan Knox, Pete Nowak, Mark Renz, Jake Vander Zanden, and two anonymous reviewers provided edits to earlier drafts of this manuscript. We would like to especially thank all the volunteers and professionals who participated in our training events in Wisconsin and Colorado.

References

Bandura, A. (1994) Self-efficacy. Pages 71–81 in V.S. Ramachaudran, editor. *Encyclopedia of human behavior*. Academic Press, New York.

Barnett, D.T., Stohlgren T.J., Jarnevich C.S. *et al*. (2007) The art and science of weed mapping. *Environ Monitor Assess* **132**, 235–252.

Bell, S., Marzano M., Cent J. *et al*. (2008) What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. *Biodivers Conserv* **17**, 3443–3454.

Bloniarz, D.V., Ryan H.D.P. (1996) The use of volunteer initiatives in conducting urban forest resource inventories. *J Arboric* **22**, 75–82.

Bonney, R., Cooper C.B., Dickinson J. *et al*. (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy. *Bioscience* **59**, 977–984.

Brandon, A., Spyreas G., Molano-Flores B., Carroll C., Ellis J. (2003) Can volunteers provide reliable data for forest vegetation surveys? *Nat Areas J* **23**, 254–261.

Brossard, D., Lewenstein B., Bonney R. (2005) Scientific knowledge and attitude change: the impact of a citizen science project. *Int J Sci Educ* **27**, 1099–1121.

Brown, W.T., Krasny M.E., Schoch N. (2001) Volunteer monitoring of nonindigenous invasive plant species in the Adirondack Park, New York, USA. *Nat Areas J* **21**, 189–196.

Campbell, D.T., Stanley J.C. (1963) *Experimental and Quasi-experimental designs for research*. Houghton Mifflin Company, Boston.

Cooper, C.B., Hochachka W.M., Dhondt A.A. (2007) Contrasting natural experiments confirm competition between house finches and house sparrows. *Ecology* **88**, 864–870.

Crall, A.W., Newman G.J., Jarnevich C., Stohlgren T.J., Waller D.M., Graham J. (2010) Improving and integrating data on invasive species collected by citizen scientists. *Biol Invasions* **12**, 3419–3428.

Danielsen, F., Burgess N.D., Balmford A. (2005) Monitoring matters: examining the potential of locally-based approaches. *Biodivers Conserv* **14**, 2507–2542.

Delaney, D.G., Sperling C.D., Adams C.S., Leung B. (2008) Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biol Invasions* **10**, 117–128.

Engel, S.R., Voshell J.R. (2002) Volunteer biological monitoring: can it accurately assess the ecological condition of streams? *Am Entomol* **48**, 164–177.

Friedman A.J., editor. (2008) *Framework for evaluating impacts of informal science education projects (Online)*. The National Science Foundation. Available from: http://insci.org/resources/Eval_Framework.pdf

Genet, K.S., Sargent L.G. (2003) Evaluation of methods and data quality from a volunteer-based amphihian call survey. *Wildl Soc Bull* **31**, 703–714.

Jones, J.P.G., Andriamarovololona M.M., Hockley N., Gibbons J.M., Milner-Gulland E.J. (2008) Testing the use of interviews as a tool for monitoring trends in the harvesting of wild species. *J Appl Ecol* **45**, 1205–1212.

Judge, T.A., Jackson C.L., Shaw J.C., Scott B.A., Rich B.L. (2007) Self-efficacy and work-related performance: the integral role of individual differences. *J Appl Psychol* **92**, 107–127.

Kennedy, K.A., Addison P.A. (1987) Some considerations for the use of visual estimates of plant cover in biomonitoring. *J Ecol* **75**, 151–157.

Krippendorff, K. (2004) *Content analysis: an introduction to its methodology*. SAGE Publications, Thousand Oaks, CA.

Lacy, S., Riffe D. (1996) Sampling error and selecting intercoder reliability samples for nominal content categories: sins of omission and comission in mass communication quantitative research. *Journal Mass Commun Q* **73**, 969–973.

Lepczyk, C.A., Boyle O.D., Vargo T.L. *et al*. (2009) Symposium 18: citizen science in ecology: the intersection of research and education. *Bull Ecol Soc Am* **90**, 308–317.

Main, M.B. (2004) Mobilizing grass-roots conservation education: the Florida master naturalist program. *Conserv Biol* **18**, 11–16.

Marcinkowski, T. (1997) Assessment in environmental education. Pages 143–197 in R.J. Wilke, editor. *Environmental education teacher resource handbook: a practical guide for K-12 environmental education.* Corwin Press, Inc., Thousand Oaks, CA.

McLaren, A.A., Cadman M.D. (1999) Can novice volunteers provide credible data for bird surveys requiring song identification? *J Field Ornith* **70**, 481–490.

Mumby, P.A., Harborne A.R., Raines R.P., Ridley J.M. (1995)A critical appraisal of data derived from Coral Cay conservation volunteers. *Bull Mar Sci* **56**, 737–751.

National Science Board. (1996) Science and technology: public attitudes and public understanding. Pages 7-1–7-21 in National Science Board, editors. *Science and engineering indicators-1996*. US Government Printing Office, Washington, DC.

Nerbonne, J.F., Ward B., Ollila A., Williams M., Vondracek B. (2008) Effect of sampling protocol and volunteer bias when sampling for macroinvertebrates. *J North Am Benthol Soc* **27**, 640–646.

Nickell, G.S., Pinto J.N. (1986) The computer attitude scale. *Comput Hum Behav* **2**, 301–306.

Penrose, D., Call S.M. (1995) Volunteer monitoring of benthic macroinvertebrates – regulatory biologists perspectives. *J North Am Benthol Soc* **14**, 203–209.

Rist, J., Milner-Gulland E.J., Cowlishaw G., Rowcliffe M. (2010) Hunter reporting of catch per unit effort as a monitoring tool in a bushmeat-harvesting system. *Conserv Biol* **24**, 489–499.

Stern, P.C., Dietz T., Guagnano G.A. (1995) The new ecological paradigm in social-psychological context. *Environ Behav* **27**, 723–743.

Stevenson, R.D., Haber W.A., Morris R.A. (2003) Electronic field guides and user communities in the eco-informatics revolution. *Conserv Ecol* **7**, 3.

Stohlgren, T.J. (2007) *Measuring plant diversity: lessons from the field*. Oxford University Press, New York.

Tonterri, T. (1990) Inter-observer variation in forest vegetation cover assessments. *Silva Fennica* **24**, 189–196.

Usher, E.L., Pajares F. (2008) Sources of self-efficacy in school: critical review of the literature and future directions. *Rev Educ Res* **78**, 751–796.

Williams, K., Sader S.A., Pryor C., Reed F. (2006) Application of geospatial technology to monitor forest legacy conservation easements. *J For* **104**, 89–93.

