Training Citizen Scientists for Data Reliability: A Multiple Case Study to Identify Themes in Current Training Initiatives

M. Gaddis
11/30/2018
Introduction

• Citizen scientists are volunteers who participate in scientific activities under the guidance of professional scientists and organizations

• This dissertation inquiry investigated citizen science training and its perceived effectiveness
Statement of the Problem

The general problem is that some scientists and land managers view the data collected by citizen scientists as unreliable, which results in lower confidence in and scientific use of citizen scientist-collected data.

The specific problem is the absence of educational training measurement in citizen science program design and analysis with which to ascertain the learning gains of trained citizen scientists.
Purpose of the Study

The purpose of this qualitative comparative multiple case study was to identify patterns and themes in content, instructional design, theoretical alignment, and perceived efficacy of training for citizen scientists tasked with collecting ecological data in the field.
Importance of the Study

- Citizen scientists are increasingly important to the implementation and on-going assessment of
  - ecological restoration
  - species identification
  - ecological monitoring on natural lands
  (Bonney et al., 2014; Dickinson, Zuckerberg, & Bonter, 2010; Handel, Saito, & Takeuchi, 2013; Havlick et al., 2014; Maschinski, Wright, & Lewis, 2012; Schmeller et al., 2008)

- Myriad other scientific contexts
Definition of Terms

• **Andragogy**: “art and science of helping adults learn” (Knowles, 1980, p. 43)

• **Backwards design**: a principle of instructional design in which the learning outcomes are identified first, followed by the design of the assessments, and then the learning activities (Wiggins & McTighe, 2005)

• **Citizen science**: “In North America, citizen science typically refers to research collaborations between scientists and volunteers, particularly (but not exclusively) to expand opportunities for scientific data collection and to provide access to scientific information for community members” (Citizen Science Central - The Cornell Lab of Ornithology, 2017)
• **Citizen scientists**: individuals who participate in scientific activities under the guidance of professional scientists and organizations

• **Informal science education (ISE)**: a term used by the United States government to refer to citizen science activities and other science learning experiences that occur outside of traditional academic settings (Bonney et al., 2009a)

• **Science literacy**: “the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity. It also includes specific types of abilities” (National Research Council, 1996, p. 22)
The literature revealed three foci of scholarly citizen science inquiry:

- science literacy gains for volunteers who participate in citizen science
  (Crall et al., 2013; Havlick et al., 2014; Lynch, 2016; Rasmussen, 2015)

- credibility of data collected by citizen scientists
  (reviewed in Dickinson, Zuckerberg, & Bonter, 2010; Kremen, Ullman, & Thorp, 2010; Reynolds, 2016; Storey & Wright-Stow, 2017)

- the usefulness and economic benefits of volunteers to natural resource management and monitoring
  (Gollan, De Bryun, Reid, & Wilkie, 2012; Handel et al., 2013)
Figure 1. ProQuest literature returns for citizen science with no date restrictions
**Citizen Scientist Demographics and Types of Citizen Science Data Collected**

- **Citizen science engages:**
  - indigenous human populations  
    (Baker, 2013; Dolrenry, Hazzah, & Frank, 2016)
  - college students  
    (Davis, Belote, Williamson, Larson, & Esch, 2016; Oberhauser & LeBuhn, 2012)
  - students of all ages  
    (Doran & Montmerle, 2012)
  - private land owners  
    (Dickinson, Zuckerberg, & Bonter, 2010)
  - non-profit groups  
    (Clarridge, 2016)
  - scientific researchers

- **Citizen science occurs in:**
  - forests  
    (Danielsen et al., 2014; Davis, Belote, Williamson, Larson, & Esch, 2016; Toman & Shindler, 2006)
  - gardens  
    (Birkin & Goulson, 2015)
  - oceans  
    (Koss et al., 2009; Van der Velde et al., 2017)
  - rivers  
    (Kruse, 2014; Storey & Wright-Stow, 2017)
  - the sky  
    (Doran & Montmerle, 2012; Henden, 2011)
  - from within one’s own home or office with the use of computer-accessible cloud-based data  
    (Ricci, 2015)
Results of the Training Literature Review

- 5% of the citizen science literature references training
- Over 150 scholarly documents analyzed for this literature review
- The journal *Citizen Science: Theory and Practice* is the first peer-reviewed journal dedicated to reporting citizen science-related research
  - first issue published in 2016
  - 6 articles from this journal are cited in this literature review (2 addressed training)
  - 10% of all articles in *Citizen Science: Theory and Practice* addressed training
Science Literacy Gains as a Result of Participation

- Citizen science projects are rarely started, if ever, for the sole purpose of enhancing science literacy.

Photo credit: http://www.openscientist.org/2013/01/the-levels-of-citizen-science.html
• Citizen science has the potential to increase the scale of data collection, but data error and bias are poorly understood (Dickinson et al., 2010)

• Discussions of data quality are the most numerous discussions in the literature currently (Clarridge, 2016; Crall et al., 2011; Danielsen et al., 2014; Gollan, De Bryun, Reid, & Wilkie, 2012; Jordan, Sorenson, & Ladeau, 2017; Kremen, Ullman, & Thorp, 2010; Koss et al., 2009; Moyer-Homer, Smith, & Belt, 2012; Reynolds, 2016; Schmeller et al., 2008; Storey & Wright-Stow, 2017)
Training is Needed

• Training and learning in citizen science have not been reported in the literature or been evaluated as rigorously as the direct comparisons of data collection efforts.

• A collection of papers support a simple conclusion that training is valuable regardless of its design.
  
  (Gallo and Waitt, 2011, Fuccillo, Crimmins, De Rivera, and Elder, 2015), Van Horn, Zug, LaCombe, Velez-Liendo, & Paisley, 2014)
Training is Needed

- All studies that address training do so in one of two ways
  - a simple fact of programming
  - training is the premise of the research
Investigations of Training Modes

• Crall and Newman (numerous publications) found that citizen scientists
  • are proficient at collecting simple data, but their data quality decreases
    with increasing data collection challenge
  • perform equally well with live or video training, and better than static
    documents alone
  • struggle with procedural and GPS skills

• Starr et al. (2014) found that
  • perform equally well with live or video training, and better than static
    documents alone
The purpose of this research design was to identify patterns and themes in content, instructional design, theoretical alignment, and perceived efficacy of training for citizen scientist volunteers tasked with collecting data in the field.
Theoretical Framework

- Educational experiences contribute to learning.
- The acquisition of procedural knowledge is possible in a training setting outside of academia where scientists are formally trained.
- Andragogy and backwards design are learning and design principles that are successful in other education and training contexts (Mitchell & Sonora, 2012; Salas et al., 2012; Smith, 2015; Toman & Shindler, 2006).
Research Questions

- R1: What are the characteristics of citizen science trainings designed to train volunteers to collect ecological data in natural land settings?

- R2: How do organizational leaders describe their perception of the efficacy of the trainings to produce reliable data collection?
Assumptions

• The whole phenomenon of citizen science assumes
  • Individuals lacking a terminal degree in science can contribute to the collection of scientific data
  • Procedural knowledge may be acquired in a training environment
  • Education experience contributes to learning
Overview of Research Design

• A qualitative multiple case study design
  • appropriate for this inquiry because it afforded a breadth of information from which themes can be extrapolated
  • no observational study currently exists to characterize citizen science training yet scholars are debating the quality of the data collected by citizen scientists (Danielsen et al., 2014; Smith, 2015)

• The research design was exploratory in nature, but the sequential investigative steps drove the investigation into the realm of descriptive case study design (Yin, 2013)
Methodology

• Involved the collection of both text-based and numeric data into one analytical tool

• Sequential investigation
  • case identification
  • training document analysis
  • organizational survey
  • follow-up semi-structured interviews with training leaders
Participant Sample

• Purposive sample
  • organizational users of the CitSci.org database, who task volunteers with field-based ecological data collection

• CitSci.org
  • free for individuals and organizations to use
to conduct citizen science of any kind
  • 400 organizational users
Data Saturation

• Goal: 15-36 cases in the case study database (Bertaux, 1981; Mason, 2010)
  • Final discussion represents 22 cases

• Are we there yet?
  • Yes, but
  • more cases requesting participation
Data Analysis Process

From the CitSci.org database

- Case Identification
- Training Document Analysis
- Organizational Survey
- Interviews with Training Leaders

Data collected in the early phases will inform the semi-structured interviews as the interviews are intended to corroborate and/or clarify training themes already identified.

Triangulation of qualitative data promotes confidence in identified themes.
Trustworthiness

• Multiple streams of analogous data
  • facilitated the establishment of qualitative data quality
• created a redundant inquiry system in which codes and themes revealed were confirmed or contradicted
  • Confirmation - demonstrated data credibility, sequence of data collection confirmed emerging themes
  • Contradiction - one exception: the discussion versus delivery of evaluation artifacts
• Codes and themes applied to a several cases
  • Demonstrated transferability and dependability
Who Manages Citizen Science Projects?

A majority of the citizen science programs in the multiple case study commenced from a single concerned person or a small group of concerned people addressing ecological issues in their communities.
Document Analysis

• Deductive codes → framework for analyzing these data
• This process of pre-determined deductive codes to which the data are applied is known as directed content analysis (Hsieh & Shannon, 2005)
• Training media delivered came in many formats
  • web links
  • static documents
  • videos
  • presentations
• 83 training artifacts
• All media types represented
• Evaluation documents under-represented
• *train the trainer*
• field engagements not well-represented in document analysis
Citizen Science Context

Types of Resources Monitored Citizen Science Programs in This Multiple Case Study

Ecosystems in Which Case Study Programs Engage Citizen Scientists

- 37% terrestrial
- 37% riparian
- 26% marine
Semi-structured Interviews and Survey Data

The survey and semi-structured interview data shed light on the perception of data reliability among citizen science training leaders.

- Demographics
- Program Statistics
- Data Reliability
- Description of QA/QC processes

- Training characterization
  - Like a traditional science class
  - Lecture
  - Hands-on component in the
    - classroom and/or
    - in the field
- Field engagements
  - mentoring
  - social learning
  - ad hoc evaluation
Instructional Design Themes

- The literature review indicated that the inherent learning outcome in most citizen science programs was likely *after training, citizen scientists will learn how to collect reliable data.*
  - All cases iterated this assumed learning outcome
- Evidence of backwards design was not found
- Only 16% of the training documents processed in the document analysis related to evaluation
- Relied on ad-hoc training
Bimodal Training Characterization

- Live training
- Complex datasheets
- Refresher trainings
- QA/QC

- Simple tutorials
- Asynchronous onboarding
- Little training burden
- Research grade data
Do you think your citizen scientists are collecting reliable data?

• 100% of survey respondents replied “yes”

How reliable do you think the data is that your citizen scientists collect?

• 82% ranked 6 of 7 in the Likert scale

• Semi-structured interview data confirmed these responses
Data Reliability

Some studies from the literature review indicated that citizen scientists are successful at collecting reliable data when tasked with a relatively simple data collection task (Smith, 2015, Crall et al., 2011, Kremen, Ullman, & Thorp, 2010, Koss et al., 2009).

• QA/QC effort
  • Ex. near 100% accuracy in the citizen science data
    • Among hundreds of data entries, there were only 2 errors that were tracked to equipment issues
  • The two other cases that employed QA/QC processes reported quantitative data reliability of over 80%.

→ These remarkable data reliability results are not reported in the literature
Do you think the data collected by your citizen scientists could be used in scientific analysis? What leads you to this conclusion?

• “I do think so, but some measures are inherently going to have more error than other. The temperature of the stream and air are very reliable, but there is more variance in taxa richness because it depends on so many factors. Even though everyone is taught the same protocol there is observer bias.”

This comment supports the literature conclusions that the reliability of citizen scientist-collected data may be related to the skill required to perform the data collection task (Danielsen et al., 2014; Schmeller et al., 2008).
Final Remarks - Data Reliability

• Categorical judgments about data reliability in citizen science are inappropriate
  • the demographics of citizen scientists do not matter to their data collection success (Danielsen et al., 2014; Schmeller et al., 2008)
  • Schmeller et al. (2008), meta-analysis of monitoring protocols in five European countries to evaluate the assumption that citizen scientists cannot collect quality data
    • Concluded that the quality of citizen science data is more likely determined by survey design, analytical methodology, and communication skill than it is a function of participant demography.

This dissertation suggests that the identity of citizen science is less relevant to the probability of data collection success than are the training methods or resource monitored.
What Citizen Scientists Do

One respondent provided the following summation of citizen scientists’ role in scientific research:

- "Site specific and regional analysis led to the conclusion that this data is analogous to what has been collected by professional scientists and in partnership with the state university. The scientists are doing research and they are working with our volunteers to expand their data collection capacities."
Theoretical Alignment: Can Training Impart Knowledge?

- Data collected in this analysis validates the deductive assumption that training can impart the knowledge required for citizen science participation.

- To put it in a different light, this multiple case study lends evidence that citizen scientists who, by definition, do not have terminal degrees in science can contribute to scientific enterprise with data collection-appropriate procedural training.
Theoretical Alignment- Adult Education

• Andragogy:
  • Activate adults’ prior knowledge, of which they have a great deal as compared to child learners.
  • Apply this knowledge to their own world (Merrill, 2002).

  • Some respondents described the practicality of knowing how to collect the data for the citizen scientist role, but more described a cultural phenomenon that the scholarly literature portrays as science literacy.

• Describe how the new learning from training is applied to the citizen scientists’ world?
  • Every respondent described increases in science literacy and a radiation of the conservation issue to the broader community through the people who participated in the trainings and subsequently collected data.

• Public Participation in Scientific Research (Shirk et al., 2012)
  • The original funding (PPSR) from which the American citizen science movement blossomed met its goal of increasing science literacy.
How do you motivate citizen scientists to persist in their role?

- Intrinsic motivation
- Social engagement

- Could an adult be motivated by bagel?
Activate Prior Knowledge

How do you activate citizen scientists’ prior knowledge in training?

• Social learning
• Life experience
• No abstract responses
  • Math skills
“All of this is real world.”

- 350,000 invertebrates tagged and many recovered in another country
- Cessation of natural resource extraction from water temperature and clarity measurements
- Identification of a little known shark as big as a great white, identified and tracked around the globe
Evaluation

Under-represented in the investigation
• Area for opportunity

Formal evaluative components
• Build credibility
• Just-in time mentoring
  • Curriculum improvement
  • Data reliability
  • Facilitate meta-analysis of training

Percentages of Documents Supporting Each Stage of Citizen Scientist Engagement
- Training: 86%
- Data collection: 23%
- Evaluation: 16%
Continuum of Citizen Science Engagement

- Most cases represented contributory citizen science
- Most hired citizen science managers did not participate in collegial/collaborative citizen science
Research Process Reflection

• The sequential design of the study was in itself adherent to principles of learning.
• The directed content analysis was informed by the deductive insights from the literature, but left room for exploring unpredicted trends and issues.
• The survey and semi-structured interview data revealed that cases:
  • Represented a two to 32-year range of citizen science engagement.
  • Involved 6-15,000 citizen scientists per year.

  • A larger sample size could shed light on the interactions between organization type, program size, duration and other nuances of the training design and data reliability outcomes.

Things to consider:
• Utility of the survey.
• Sequence of the data collection.
• Power with a survey tool in a multiple case study.
Summary

• Qualitative, comparative, multiple case study design
• Purposive sample - CitSci.org database
• Sequential design: case identification, document analysis, survey, semi-structured interviews
• Cross-case analysis
  • Revealed themes across citizen science training programs
  • Revealed demographic and program trends across the citizen science landscape
    • Programs are run by non-profit organizations
    • Started by concerned citizens
    • Sometimes supported with university and government partnership
    • Citizen scientists represent a broad array of demographic characteristics
• Bimodal training characterization:
  • Tutorials support photo data collection
  • Live trainings support complex measurements
• Evaluation is an opportunity for improvement across citizen science programs
Post-Script: Why isn’t this Grounded Theory Research?

A question remained until nearly the final hour—why isn’t this grounded theory research? I asked this question many times...

Answers included:
“citizen science is not a phenomenon.” I beg to differ...
“grounded theory is about emotions.” That’s not really true because *phenomena* can take myriad forms, not just emotional response.

Here is my best answer:

- It comes down to the theoretical foundation of the research. As I was writing the final chapters, I wanted to emphasize that the study design was informed by the literature. This is a deductive approach. The context already existed in my mind. The data were “tested” against the codes arising from the literature.

- Grounded theory research is based on an inductive perspective in which the characterization of the phenomenon arises from a pure (and ignorant?) mental state that is open for interpretation because it is not sullied with what it already known. I am thinking of Jane Goodall’s early days as a great example of grounded theory research.

- So, not grounded theory. Multiple Case Study research. Tell your future students. 😊
References


References

References


Acknowledgements

• Thank you committee
• Thank you family