Evaluating participant outcomes in the weather monitoring citizen science project “Delft meet regen”

Janssen, J.; Bogert, M; Gold, M.; Vries, S. de

CITIZEN SCIENCE:
THEORY AND PRACTICE

Evaluating Participant Outcomes in the Weather Monitoring Citizen Science Project "Delft Meet Regen"

RESEARCH PAPER

JORDY JANSSEN
MARIT BOGERT
MARGARET GOLD
SANDRA DE VRIES

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

ubiquity press

ABSTRACT

Understanding the outcomes and impacts achieved in citizen science (CS) initiatives is essential for evaluating and enhancing participatory research practices. There is increasing attention paid in the literature to the broad range of benefits experienced by participants in CS initiatives. While existing frameworks provide structured guidance for the evaluation of participant outcomes, they might overlook nuanced participant experiences. This study explores outcomes reported by participants from the Delft Meet Regen (Delft Measures Rain, DMR) project, a CS initiative focused on environmental monitoring. Using semi-structured qualitative interviews to inductively collect participant outcomes, we applied a combination of the Individual Learning Outcomes (ILO) and Impact Domain (ID) frameworks to reflect on these outcomes with participants. Participants reported meaningful experiences such as a sense of contribution, increased awareness of rainfall patterns, and strengthened feelings of pre-existing environmental stewardship. These outcomes did not neatly align with predefined categories of the ILO and ID frameworks, and this illustrates the complexity of evaluating participant outcomes. Our findings highlight the value of integrating inductive qualitative methods within structured evaluations. By sharing these findings, we aim to support the continued development of comprehensive evaluation tools that fully capture the complex experiences of citizen scientists.

CORRESPONDING AUTHOR:
Jordy Janssen
Citizen Science Lab (Leiden University), NL
janssenj4@vuw.leidenuniv.nl

KEYWORDS:
citizen science; participant; outcomes; impact; evaluation; framework

TO CITE THIS ARTICLE:
Janssen, J., Bogert, M., Gold, M. and de Vries, S. 2025. Evaluating Participant Outcomes in the Weather Monitoring Citizen Science Project "Delft Meet Regen". *Citizen Science: Theory and Practice,* 10(1): 20, pp. 1–14. DOI: https://doi.org/10.5334/cstp.757

INTRODUCTION

OUTCOMES OF CITIZEN SCIENCE

Citizen science (CS) produces new knowledge in collaboration with actors outside the academic or professional research community. This approach is now well recognized in the literature and has been defined numerous times within different domains and contexts (Auerbach et al. 2019; Heigl et al. 2019; Haklay et al. 2021). At the start of the century, reports on CS practices focused primarily on the ability to collect a large variety and quantity, as well as high resolution of data. Increasingly, other benefits such as contributions to research quality (Finger et al. 2023) or to the CS participants themselves have received more attention (Phillips et al. 2018; Shirk and Bonney 2018; Peter et al. 2021).

Participation in CS projects can result in learning benefits regarding the scientific process and the specific domain under investigation (Jordan et al. 2011; Peter, Diekötter, and Kremer 2019; von Gönner et al. 2023). Participation can also lead to changes in attitudes or behaviours towards the topic being investigated (Brossard, Lewenstein, and Bonney 2005; Turaga, Howarth, and Borsuk 2010; Jordan, Ballard, and Phillips 2012).

In the case of environmental monitoring CS initiatives, participants are able to learn more about environmental issues and might feel more connected with nature, which can strengthen their environmental citizenship (Jørgensen and Jørgensen 2021), and encourage a positive outlook on their ability to make an impact on these issues (von Gönner et al. 2023). The potential of these latter benefits seems clear, yet they can be difficult to measure. Additionally, it is not yet evident that this potential is consistently delivered on. Efforts to improve and standardise the evaluation of CS have resulted in the creation of tools such as evaluation frameworks and toolkits that offer guidance for CS practitioners who want to measure the impacts of their projects (Phillips et al. 2017; Somerwill and Wehn 2022; Finger et al. 2023; Land-Zandstra et al. 2023).

THEORETICAL FRAMEWORKS FOR EVALUATING THE OUTCOMES AND IMPACTS OF CITIZEN SCIENCE

To investigate whether a specific CS project has met its desired objectives, evaluators differentiate among the terms outputs, outcomes, and impacts. These are based on a logic model of evaluation (Örtengren 2004; Schäfer et al. 2021). As defined by Wehn et al. (2021), outputs are the products of a CS initiative, such as data or reports; outcomes refer to resulting changes over the short-term, such as changes in people's behaviour; and impacts refer to changes over the long-term, such as improved policy or environmental progress.

To facilitate the evaluation of CS projects, Kieslinger et al. (2018) developed a framework that recommends the evaluation of outcomes and impacts along three dimensions: scientific impact, learning and engagement of participants, and impact for wider society. A thorough review conducted by Wehn et al. (2021) of existing impact studies in the literature identified a total of five domains: Society, Economy, Environment, Science & Technology, and Governance. They further developed these into an Impact Domain (ID) framework and noted that the domains and their outcomes are not distinct, but often overlap.

Building on the field of Informal Science Education, Phillips et al. (2018) developed a framework describing outcomes within the participant dimension in more detail through a comprehensive study of CS projects' intended learning outcomes. This framework describes six Individual Learning Outcomes (ILOs): Interest for Science and the Environment; Motivation for Science and the Environment; Self-efficacy for Science and the Environment; Knowledge of the Content of, Process of, or the Nature of Science; Skills of Science Inquiry; and Behaviour and Stewardship (Figure 1).

QUALITATIVE EVALUATION METHODS FOR CITIZEN SCIENCE OUTCOMES AND IMPACTS

Such frameworks are valuable tools to help practitioners distil research outcomes, observations, or survey results into measurable outcome categories. Quantitative research is a common method for surveying participants, which uses predetermined questions often based on existing frameworks or previously published outcomes (Mohajan 2018). Qualitative methods such as interviews or narrative storytelling are more appropriate in exploratory research (Constant and Roberts 2017), where the answers might require time to understand, the outcomes might not be commonly defined (Jensen and Laurie 2016; Mohajan 2018), and more contextual data is provided (Lynch et al. 2018). These methods allow participants to express more nuanced experiences, which might get lost when distilling outcomes into frameworks.

This concern is supported by Peter, Diekötter, and Kremer (2019), who conducted a systematic review of biodiversity CS projects and highlighted the need for further research into participant outcomes, particularly those that go beyond changes in knowledge. They also discussed several outcomes that were valued by participants that do not fit the ILO framework, such as an increased appreciation for their environment, a sense of satisfaction and contribution, health and wellbeing benefits, and a negative effect of pessimism about the future of biodiversity. They concluded that future research should investigate participant outcomes more deeply and systematically across various projects. Other studies support this view, highlighting a

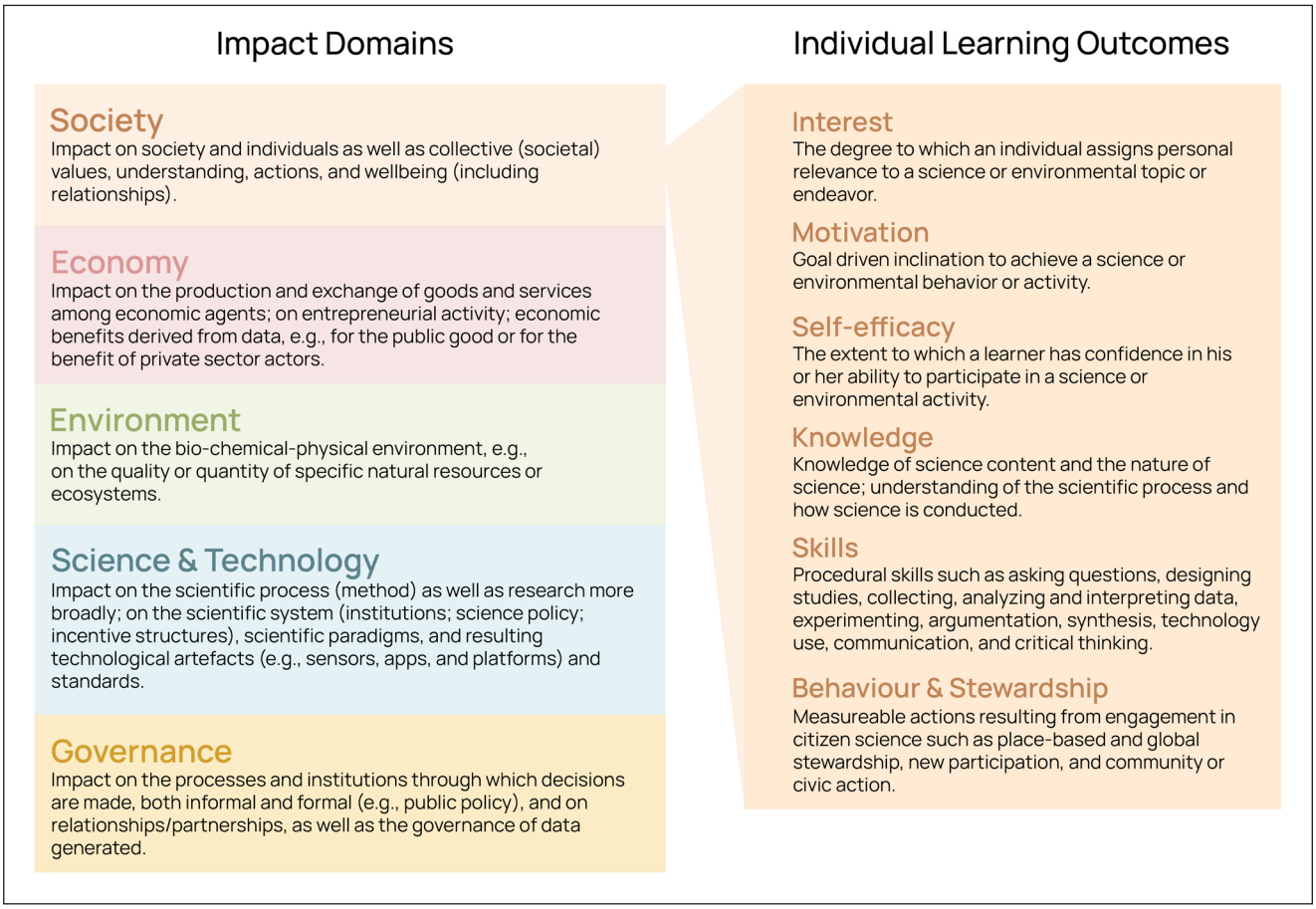

Figure 1 A visual representation of the Individual Learning Outcome framework (Phillips et al. 2018) as part of the Society domain from the Impact Domain framework (Wehn et al. 2021), with definitions.

similar gap between typically evaluated outcomes and the more personal, nonrational outcomes of participant experiences (Bela et al. 2016; Bonney et al. 2016; Groulx et al. 2017). This latter category of outcomes could reflect long-term, transformative impacts on participants, which remain relatively underreported and are not easily evaluated in a standardised manner. As a result, conclusions about participant impacts in CS projects remain difficult to draw (Land-Zandstra et al. 2021; Schaefer et al. 2021).

AIMS OF THE STUDY: QUALITATIVE EVALUATION OF DELFT MEET REGEN

As organisers of the CS project Delft Meet Regen (Delft Measures Rain, DMR), we often evaluated our project using quantitative surveys. However, we suspected that our approach overlooked part of the full range of participant outcomes. In keeping with the third principle of CS as formulated by the European Citizen Science Association (ECSA) that "both the professional scientists and the citizen scientists benefit from taking part" (Robinson et al. 2018), this study aimed to better understand the outcomes of the DMR project as experienced by its participants. Such understanding is valuable for improving project design, sustaining engagement, and ensuring that participants benefit from taking part. To capture this participant perspective, we began by inductively gathering self-reported outcomes before introducing terms from CS evaluation frameworks. This formed the basis for an explorative method that combines the structure provided by evaluation frameworks with the openness of qualitative interviews. Specifically, we applied both the high-level ID framework and the more detailed participant-level ILO framework. These frameworks overlap conceptually, especially in the Society domain, which is defined as both individual and collective level outcomes, or changes in values, attitudes, relationships, and community dynamics (see Figure 1). We expected this combination to sufficiently cover most outcomes gathered by our study. Following Wehn et al. (2021), we hypothesised that comparing participant-reported outcomes to these frameworks would yield useful insights for improving future evaluation approaches. Therefore, during our semi-structured interviews, we encouraged interviewees to reflect on their reported outcomes after being introduced to these frameworks.

The research questions identified for this study were: (1) What outcomes have participants of DMR experienced during and after the project?, and (2) How do these outcomes align to known outcomes of CS projects as represented in evaluation frameworks such as the ILO and ID framework?. In the following sections, we present our methods and findings before discussing the value of qualitative methods for identifying a broader range of outcomes.

METHODS

THE DELFT MEET REGEN PROJECT

DMR is a multi-year, city-based CS project initiated in 2020 and executed by the water- and weather-focused CS platform WaterLab. Residents of Delft were invited to help scientists at the Delft University of Technology (TU Delft) study how the city's layout locally influences rainfall, by measuring local weather and soil moisture in their own gardens with a provided rain gauge. The invitation was extended again in the summer of 2021, and 105 residents participated in that second edition. This group was approached in February of 2022 for participation in the study discussed herein. At the time of writing, the project continues to expand, now including a digital weather station and soil sensors, with more than 8,500 measurements by citizen scientists to date. The participants' measurements provide a more detailed view of varying rainfall from neighbourhood to neighbourhood, along with contextual data about green-cover. This expands the data coverage beyond the Royal Netherlands Meteorological Institute (KNMI) average of one weather station per 100 square kilometres throughout the Netherlands to an aimed-for coverage of 1 weather station per 1 square kilometre in Delft. All results and research reports are openly available on the project website (WaterLab 2024).

DATA COLLECTION

We performed semi-structured, qualitative interviews with the DMR participants to understand what outcomes they experienced before introducing them to any frameworks (Figure 2). The interviews consisted of three parts, with each

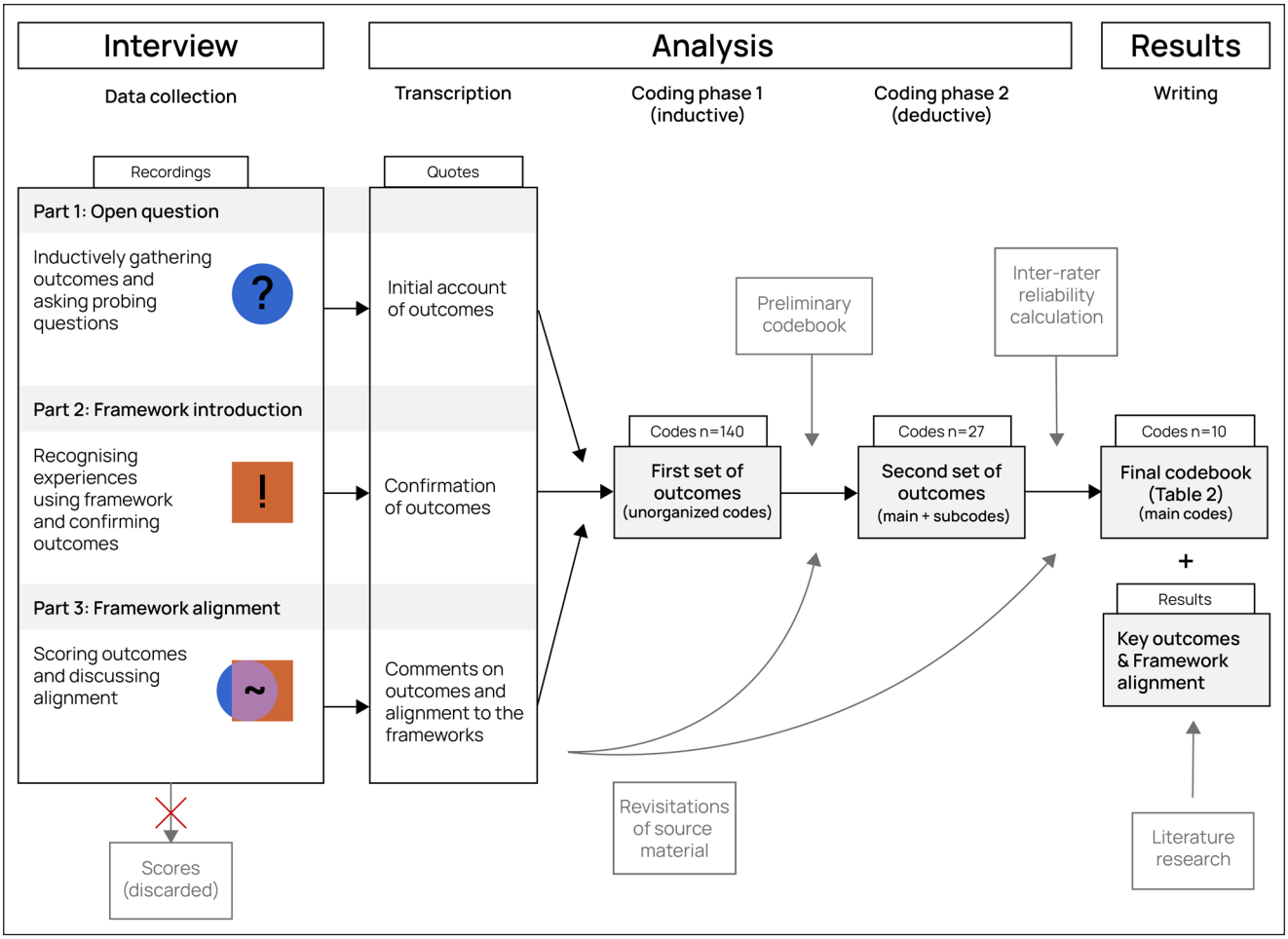

Figure 2 Visual overview of the research methods, showing how quotes from the three parts of the interview were coded in two phases, resulting in the final codebook.

subsequent part being less inductive and focusing more on the frameworks: (1) opening question and responses; (2) framework introduction and specifying outcomes; and (3) mapping outcomes on the framework. An interview guide was created to structure questions or reminders according to the interview schedule (Supplemental file 1: Interview Guide). This three-stage approach to the interviews helped to collect a variety of comments on the outcomes and frameworks from the interviewees, but was treated as one interview during analysis.

All 105 participants in the 2021 edition of the DMR project were invited by email to take part in these interviews. A few interviewees (n = 2) replied via email but the majority (n = 14) was recruited by calling a randomised list of the remaining participants. We stopped recruitment at a feasible number of interviewees given our resources. The first interview was held with a participant who was a professional in CS and water management at the TU Delft, which was used as a test interview. The interview duration ranged from 25 to 75 minutes, depending on the conciseness of interviewees, how many outcomes they identified, and their willingness to enter into a longer discussion of their experiences. All interviews were conducted by the first author in Dutch, except for one interview with a participant who could understand Dutch but was more comfortable in English.

After an introduction to cover data-ethics best practices and establish rapport, part 1 started with the open question: "Could you tell me in what ways participating in DMR has been meaningful for you, in terms of the outcomes or results that you experienced from the project?". Probing questions were asked to help interviewees elaborate on their answers, and reminders of key aspects of the project (such as the measuring task, or information received via newsletters and webinars) were given by the interviewer only after the first top-of-mind answers were noted.

For part 2 of the interviews, a physical two-sided board was created that displayed the frameworks used in this study (Supplemental file 2: Framework Board). It was used to introduce the interviewees to the concept of evaluation frameworks by explaining the outcomes and impacts one by one. As soon as the interviewees were familiar with the frameworks, they were asked whether they experienced any of these outcomes. After any clarification and structuring of the discussed outcomes, the final list of outcomes was agreed on before continuing to part 3.

In part 3, the interviewees were involved in the process of classification of the outcomes. The interviewees were asked to score the outcomes that they experienced within the frameworks presented in part 2, which provided an opportunity to also share their views about the frameworks themselves. We explored this method to map the alignment of the outcomes to the frameworks, but did not design more formal methods for quantitative analysis due to resource constraints. Quotes from all three parts of the interview were used in the analysis, but the scores of part 3 were discarded.

ANALYSIS

All interviews were recorded. The audio files were transcribed and imported into Atlas.ti 23 software for the coding process. A preliminary codebook based on the combined ID/ILO frameworks was created (Supplemental file 3: Preliminary Codebook). Thematic analysis of the gathered information followed the first three stages of the Framework Analysis method (Gale et al. 2013; Hackett and Strickland 2019) and took place in two phases. The first phase started with inductive or data-driven coding to collect any outcomes from the complete interview transcripts. In the second phase, we switched to deductive or concept-driven coding (Gibbs 2007; Linneberg and Korsgaard 2019) by fitting the collected set of codes in the preliminary codebook.

Phase 1 (familiarisation and inductive coding)

The first author conducted initial coding by identifying quotes relevant to the research questions. Although the preliminary codebook did inform the researcher's naming convention, the codes remained descriptive to allow them to emerge directly from the data, generating a first set of descriptive codes (n = 140). The size of the coded sections, that is, units of meaning (Campbell et al. 2013), were pieces of text ranging from 11 to 529 words, each containing the amount of information that provided enough context for a complete discussion of one outcome. This inductive start of the analysis ensured that the codes were made without relying too much on the existing literature and frameworks, allowing space for new or alternative outcomes (Linneberg and Korsgaard 2019).

Phase 2 (indexing and deductive coding)

The opposite was done during coding phase 2, when we aimed to fit each of the codes within the preliminary codebook based on the ID/ILO frameworks. The codes were scanned for meaning and placed in the main code categories from the preliminary codebook. A non-applicable (N/A) category was added to isolate codes representing outcomes that directly fit the frameworks from any new or ambiguous codes, allowing further investigation and categorising of the latter (Linneberg and Korsgaard 2019). The quotes and recordings from each code within the N/A category were revisited, and were either re-coded in one of the categories from the preliminary codebook where suitable, or assigned a newly-created category where needed. Clarifications from interviewees later on in the

interview were used to change the coding of earlier quotes when necessary. Notes of this process are integrated in the results section. This second phase compressed the first set of codes into a definitive codebook representing outcomes obtained from the interviews within this study (n = 27 codes; Supplemental file 4: Definitive Codebook). The number of interviews in which a code was mentioned by the interviewees (i.e., occurrence) was counted as a variable and added as well.

Inter-rater reliability

For measuring inter-rater reliability (IRR), 10% of the quotes were randomly selected and compiled into a document (Jensen and Laurie 2016). This document was then coded by the second author only for feedback, which indicated a rather low percent agreement (PA) of 53%. An untrained second rater not affiliated to the project was then asked to code this document, which confirmed the indication with a PA of 57.1% (Table 1). A part of the disagreement was observed within the more detailed subcodes (e.g., knowledge:science or knowledge:watermanagement) as opposed to the main codes (e.g., knowledge). The PA of the second rater increased to 66.7% when the subcodes were collapsed.

We subsequently updated the codebook by adding examples and improving the grammar of the definitions. We removed two codes that were more relevant for informing the project's organisation than the research questions, and merged two sub-level codes. A third rater coded a new set of quotes based on this improved codebook and reached agreement with the first author in a discussion. A PA was reached of 66.7% for subcodes and 83.3% for main codes, reaffirming that the main codes are more representative of the meaning of quotes. Therefore, the conclusions of this study mainly focus on those.

As PA should not be used alone when the coding task is difficult (Feng 2014), we calculated two IRR indices, which were consistent with the PA values (Table 1). We calculated (Supplemental file 5: Data) Cohen's κ in Microsoft Excel Professional Plus 2021 following instructions of Zaiontz (no date), and Krippendorf's α in the online Krippendorff›s Alpha Calculator (Marzi, Balzano, and Marchiori 2024). Older, original thresholds suggest that 0.4 to 0.6 is a fair agreement (Landis and Koch 1977), but the most recent research recommends a minimum of 0.8, even though their applicability to qualitative interviews is questioned (Campbell et al. 2013; Halpin 2024).

RESULTS

In Table 2, we present the outcomes identified by our study. The interviewees mentioned all outcomes from the ILO framework developed by Phillips et al. (2018) with the exception of the Skills of Science Inquiry category, likely due to the straightforward nature of measuring rain. The ID framework developed by Wehn et al. (2021) was harder to apply in practice, and often led to speculation by interviewees about any impacts resulting from their reported outcomes. The ID categories turned out to be too abstract and broad for the concrete outcomes gathered by this study. Interestingly, we found a number of outcomes that did not directly fit either framework, which we will discuss below in more detail in the context of similar outcomes described in literature. Interviewees reported a sense of satisfaction from contributing data to research, valuing the social encounters, and experiencing an increased awareness of levels of rainfall. Many also viewed their participation as an expression of a pre-existing stewardship. These results indicate that our selection of frameworks was not well suited for capturing all types of outcomes valued by our interviewees.

CONTRIBUTING DATA OR A FEELING OF SATISFACTION

A prominent outcome mentioned by fourteen interviewees was that participation gave them a feeling of having made a contribution. The feeling of being part of a larger whole was also a recurring topic, as illustrated by the following quote:

	RATER 2		RATER 3	
	MAIN CODES	SUBCODES	MAIN CODES	SUBCODES
Percent agreement	66.7%	57.1%	83.3%	66.7%
Cohen's κ	0.59	0.53	0.80	0.64
Krippendorf's α	0.59	0.54	0.74	0.64

Table 1 Inter-rater reliability index values. A first set of quotes were rated by rater 2. After updates to the codebook and a discussion a second set was rated by rater 3. Main codes are the higher-level outcome categories that his study focuses on, while subcodes reflected more detailed definitions from the coding process (Supplemental file 4: Definitive Codebook).

CODE	CODE NAME	OCCURRENCE (n/15 INTERVIEWS)	DEFINITION
1	Contributing data	14	Participant has the feeling of having contributed
2	Project part of stewardship	9	Participant sees participation as part of their stewardship
3	Social aspects	9	Participant actively or passively gained social encounters
4	Awareness	10	Participant noticed an increase in awareness of droughts, rainfall and general changes in the weather
5	Knowledge	12	Participant gained knowledge
6	Motivation	6	Participant gained motivation
7	Confidence (self-efficacy)	2	Participant gained confidence
8	Interest	3	Participant gained interest
9	Behaviour	3	Participant has changed their behaviour
10	Other codes	10	Participant…
10.1	*Connection*	*2*	*… gained a feeling of connectedness with their surroundings or globally with similar projects and being part of a larger whole*
10.2	*Cooperation between involved institutions*	*1*	*… observed that the project connected parties relevant to the subject of the project, sparking cooperation or sharing of knowledge*
10.3	*Experience is fun to do*	*6*	*… found participation a fun, surprising or otherwise engaging experience*

Table 2 The final codebook with the outcomes of this study (collapsed to show only the main codes).

> *"Well, the fact that you think, oh well, I am just one of the 100,000 inhabitants of Delft so I won't have that much influence. But by doing one's tiny bit, apparently, together we do have influence."*

The ways in which their contribution seemed valuable varied between participants, but contributing to science was frequently mentioned (twelve participants):

> *"Every research project that contributes to better insight is valuable. So, if I can contribute a small part to that, then I will do that with contentment. … It gives me a good feeling. A little bit, but yeah, it gives me a good feeling to be able to contribute to that."*

For some, the act of contribution itself was sufficient to have this feeling, while seven interviewees saw their contribution rather as a pathway towards helping the changing climate. Five participants mentioned doing something for other people, which we coded as a form of altruism:

> *"Actually, you are doing it for the people who are asking, I mean, I don't have that much to gain from it. Apart from that I liked doing it."*

This outcome has characteristics of the ILO Behaviour & Stewardship, but that is defined as measurable actions outside the project. A better match can be found in Haywood (2016), who describes a Sense of Satisfaction and Contribution as a lasting and more existential form of fulfilment, and Pride, discussed by Greving et al. (2023).

PARTICIPATION AS A PART OF PRE-EXISTING STEWARDSHIP

While discussing the concept of stewardship, some interviewees were eager to stress that they were environmental citizens before participating in the project. For them, participation was a way of fulfilling this stewardship:

> *"I already always did my best, so to speak. But it [stewardship] could have increased a bit by my participation in such a project."*

Six interviewees said that it was increased or reinforced, but three said that participation in the project was itself a result of their stewardship:

> *"I was already actively aware of our water usage and those kinds of things. You could say that it is the other way around, that because of that I was interested in this project…. My participation was more of an expansion, a part of, or extension of that."*

An increase in stewardship is defined by the ILO Behaviour & Stewardship, but this nuanced view could be of value to evaluators. Respondents from Chase and Levine (2018), who were asked how a CS project influenced their attitudes, reported a reinforcement of their attitudes, values, and beliefs. Others reported "no change" due to strong pre-existing values, which was an outcome similar to the nuance brought up by our interviewees. Participation thus becomes a fulfilment of pre-existing stewardship rather

than the other way around, which has characteristics of a means of self-expression, that is, the expression of participants' values through their work, described by Bruyere and Rappe (2007).

SOCIAL ASPECTS AS AN OUTCOME

Societal aspects were mentioned in nine of the interviews regarding the sharing of experiences or information from the project with other people. Seven interviewees described sharing information about the project with colleagues, neighbours, or family, and three said they used the project for educational purposes with their children (two) or with high school students (one). Although Phillips et al. (2018) describe social aspects as a motivation for participation in a CS project, it is not included in their ILO framework as a desirable outcome. Social outcomes such as interactions or new relationships with like-minded people, whether scientists or participants, and increased community engagement are a common outcome of CS projects (Groulx et al. 2017) and can be categorised under the Society ID in the Wehn et al. (2021) framework. Although the increased social contact described by our interviewees was not as long-lived, it did form a valuable outcome identified during the interviews.

AWARENESS OF THE MONITORED RESOURCE

An outcome brought up by ten participants was an increased observational awareness of the monitored resource (rainfall or the weather in the case of DMR):

> *"Now for example, it hasn't rained for a while and I would never expect it to not rain for three weeks around this time. But now that really strikes me, that that happens a lot.... But this absolutely was not the world as I would look at it, and now I see ... how much rain does or does not fall. That [outcome] alone is very interesting to me. You are more aware now."*

One interviewee suggested that the act of measuring rain every day contributed to this new awareness of rainfall and enabled them to place it in context:

> *"Now for example, it has been raining a lot for a long time that I start thinking about if that is normal or not."*

When asked, seven out of the ten interviewees explicitly said that this heightened awareness was still present at the time of the interview, 5–6 months after their participation in the project. This outcome differs from (environmental) awareness of one's own behaviour and the effect it can have on the world. It shows some similarities with the ILO category Interest, but more clearly falls under the category of Greater Awareness and Appreciation described by Haywood (2016) and the Increased Awareness of the Natural World described by Chase and Levine (2018). The latter found that half of their participants noticed a change in attitude towards the resource being monitored. Haywood (2016) also argues that a heightened awareness of the project's study object is not necessarily the same as an increased understanding or a heightened connection to nature. With regard to Interest, only four interviewees specifically mentioned reading or learning more about rainfall:

> *"Because of this project I have been looking into what drought does to our country and the soil, just a little bit more in-depth."*

OTHER OUTCOMES VALUED BY INTERVIEWEES

The catch-all "other outcomes" category collected several subcodes that are worth mentioning despite their low inter-rater reliability. Six interviewees said that the project was fun to do, which has clear value for participants and fits within the Enjoyment category described by Sickler et al. (2014). Furthermore, two interviewees noted that participation gave them a feeling of connection or engagement with their hometown (Delft) or with the world, similar to a Sense of Place (Haywood, 2014):

> *"Well, I think it is a bit of engagement. It gives me the feeling that I am a bit more engaged with what happens in Delft, a bit of connection with the city so to speak."*

Finally, we highlight a quote from an interviewee who was involved in the writing of a scientific paper related to the project (Bogert et al. 2022), which is a clear example of the ILO category Self-efficacy:

> *"Well, with writing that paper, that really changed. So, without this project I would not have participated in that, or it wouldn't even have crossed my path.... But that really entails having the confidence to join doing that."*

DISCUSSION

LIMITATIONS

Selection bias

Selection bias might have influenced these findings due to the relatively small sample size (16 interviewees), primarily

resulting from project resource constraints and participant availability. The responding participants might hold stronger opinions or sentiments about the project than the average pool of participants. As we did not collect participant sociodemographic information, we are unable to know if our sample is representative, but it is highly unlikely that the results are generalisable to a broader population.

Inter-rater reliability

The values of our inter-rater reliability ranged from 0.59 to 0.83 for the main code categories. Consistent values between PA and IRR indices indicate minimal chance agreement. High agreement values are not expected because of our small sample size, the interpretative nature of the data, and the limited training or background knowledge of additional raters (Halpin 2024). We think our IRR is satisfactory for exploratory research but remain cautious of generalizing the results (Syed and Nelson, 2015).

Considerations on outcome reliability

We aimed to reduce our dependence on frameworks by starting with an open-ended question. We observed that some interviewees had difficulty recalling specific outcomes from memory. This could introduce an availability bias towards easily recalled outcomes, especially 6 months after the end of their participation. Our qualitative design intentionally facilitated reflection, aiming to inductively surface participants' most valued outcomes. Future studies aimed at quantifying the results should employ more rigorous evaluation methods to better prevent these biases. Introducing frameworks later in the interview likely aided interviewees in recalling outcomes, thereby reducing availability bias (Tversky and Kahneman 1973; Schwarz et al. 1991). However, this approach might have influenced the authenticity of participant responses by prompting them to reflect in specific directions. Additionally, self-reporting methods like ours are acknowledged to have limitations in assessing knowledge gained or behavioural changes (Kormos and Gifford 2014).

Collaborative coding and use of frameworks

Collaboratively classifying outcomes during the third part of the interview was occasionally challenging. Despite options to select multiple or no categories, interviewees found assigning outcomes to specific ID domains challenging due to their interconnected nature (Wehn et al. 2021). We included this framework specifically to enable interviewees to discuss and categorise outcomes outside of their personal experiences on an individual level. The ILO framework, though able to capture most of the outcomes, still presented challenges for some interviewees. Our method did not explicitly mitigate this by, for example, including more elaborate explanations or a training session. These difficulties, combined with our results showing qualitative outcomes that did not fit the frameworks, highlight different perspectives from participants and practitioners on evaluation frameworks. Frameworks are valuable for designing surveys, a method also used by the DMR organisers for their ease of data collection, yet this may risk the omission of information particularly valued by participants.

OUTCOMES ARE INTERTWINED WITH MOTIVATIONS, ATTITUDES, AND ENGAGEMENT

Our evaluation of the DMR project revealed several outcomes—such as fulfilment of stewardship, social experiences, and a feeling of contribution—that reaffirm the value of qualitative methods. Interviewees reported diverse outcomes, which, although varied, are examples of the intertwined nature of motivations, attitudes, engagement, and outcomes currently being discussed in the CS evaluation field.

Peter et al. (2021) identified a similarly diverse range of outcomes, emphasising the importance of evaluating attitudinal structures. These structures extend beyond the learning outcomes traditionally aimed for in CS projects (Phillips et al. 2018). The shift in terminology from Individual Learning Outcomes to Participant Outcomes by Peter et al. (2021), while not explicitly addressed, better reflects the range of participant experiences we observed. For example, our participants reported personal experiences, such as enjoyment or social interactions. While these might be considered outputs rather than outcomes, they form lasting memories valued by participants.

Some interviewees described their participation as a way to fulfil an existing sense of stewardship rather than a means to obtain it. This suggests it might be better to ascertain stewardship separately from behaviour, given it is a more complex construct encompassing awareness, attitudes, and behaviours (Jørgensen and Jørgensen 2021). This attitude among our interviewees may have contributed to their sense of contribution. As Greving et al. (2023) note, attitudes towards engagement in a CS project can influence outcomes such as the sense of contribution or motivation after the project. This means that aspects of participation commonly evaluated as motivation or engagement can themselves evolve into outcomes valued by participants. This highlights how participant outcomes such as those identified in our study are highly complex, thus increasing the difficulty of evaluation for practitioners.

This complexity is also outlined by Phillips et al. (2019) in their Dimensions of Engagement framework, which

describes engagement as a mix of motivations across behaviour, social connection, learning, and feelings. McAteer, Flannery, and Murtagh (2021) further explore this by linking participants' motivations to outcomes, identifying four types of CS participants. These each value their own set of outcomes and have different degrees of participation, which can influence their quality of participation (Shirk et al. 2012). Considering the diversity of participants in CS project design and evaluation, for example by using a Pathways to Impact approach (Pateman and West 2023), can help CS practitioners and participants to reach the full potential of their cooperation (Bieszczad, Fochler, and Brodschneider 2023).

CONTINUING THE DEVELOPMENT OF EVALUATION FRAMEWORKS

The field of CS evaluation would benefit from the consistent use of standardised evaluation methods on a larger scale (von Gönner et al. 2023). Ideally, all outcomes could be evaluated using accessible frameworks such as that developed by Phillips et al. (2018), accompanied by their methodological guide (Phillips et al. 2014, 2017). While recognising that including nuanced, inductively obtained participant outcomes complicates CS evaluation, we argue for their preservation within frameworks. This might prove to be difficult given the necessity of other methodological considerations such as the need for embedded assessment and collaborative evaluation (Becker-Klein, Peterman, and Stylinski 2016; Davis et al. 2022). Additional considerations include the often-limited availability of resources and the impracticality of measuring every outcome for each project (Phillips et al. 2018; Somerwill and Wehn 2022; Land-Zandstra et al. 2023). We therefore recommend that future evaluation frameworks continue to adopt a modular structure, similar to the ILO framework or that of the Impact Lab Toolkit (Land-Zandstra et al. 2023). From the evaluators' stand point, presenting such a framework with a flexible difficulty level enables them to make an *à la carte* selection from a longer list of potential outcomes to evaluate based on their available resources and methodological capabilities. Evaluating many CS projects using different parts of a single, many-sizes-fits-all framework would enable the use of diverse evaluation methods while facilitating their aggregation into a comprehensive overview of impacts.

CONCLUSION

The CS field aims to evaluate projects consistently and reproducibly, with standards increasingly taking shape. However, indications that some qualitative participant outcomes remain underrepresented in frameworks often used for quantitative evaluation prompted us to evaluate our DMR CS project by integrating two frameworks in qualitative interviews and include the participant in the process. Our findings from this combined approach to evaluation highlighted both the strengths and the limitations of using existing frameworks for quantitative evaluation. While they effectively structure outcomes for evaluators, they do not fully capture the complexity and nuance of participant experiences, such as their feelings of contribution, social interactions, and increased awareness of the monitored resource. These participant outcomes have, however, already been identified and discussed in the literature and are described as being intertwined with participant motivations, attitudes, and engagement. Although this added complexity is challenging for evaluators, we found several methods that aim to address this. With this study, we aim to inform the continued development of more comprehensive, flexible, and accessible evaluation frameworks, ultimately helping practitioners to better assess and enhance the transformative impacts of their CS initiatives.

DATA ACCESSIBILITY STATEMENT

Data of inter-rater reliability calculation will be uploaded on OSF.io: DOI 10.17605/OSF.IO/4UFM3. Interviewees did not explicitly give consent for placing the complete interview transcripts online. Raw redacted transcripts and coding data can be available upon request after inquiry by the authors with the ethics committee (see below) until the 3rd of July, 2032.

SUPPLEMENTARY FILES

The Supplementary files for this article can be found as follows:

- **Supplemental File 1.** Interview Guide. DOI: https://doi.org/10.5334/cstp.757.s1
- **Supplemental File 2.** Framework Board. DOI: https://doi.org/10.5334/cstp.757.s2
- **Supplemental File 3.** Preliminary Codebook. DOI: https://doi.org/10.5334/cstp.757.s3
- **Supplemental File 4.** Definitive Codebook. DOI: https://doi.org/10.5334/cstp.757.s4
- **Supplemental File 5.** Data. DOI: https://doi.org/10.5334/cstp.757.s5

ETHICS AND CONSENT

Approval has been given by the Human Research Ethics Committee of the TU Delft on 23/03/2022 to applicant Marit Bogert.

ACKNOWLEDGEMENTS

Without the participants and especially interviewees, this research would not exist. Lots of gratitude to Sanne Romp for her thoughtful motivational and substantive feedback during her supervision. Many thanks to the Science Communication & Society department of Leiden University for their education and support, allowing me to pursue research in this discipline.

FUNDING INFORMATION

This project was funded with internal funds from the TU Delft Open Science Program, the TU Delft Faculty of Civil Engineering and Geosciences, and the municipality of Delft.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

CRediT (Contributor Roles Taxonomy): **Jordy Janssen**: conceptualisation, data curation, formal analysis, investigation, methodology, project administration, visualisation, writing—original draft, review, and editing. **Marit Bogert**: conceptualisation, formal analysis, funding acquisition, methodology, resources, supervision, writing—review and editing. **Margaret Gold**: writing—review and editing. **Sandra de Vries**: conceptualisation, funding acquisition, methodology, project administration, resources, supervision, writing–review and editing.

AUTHOR AFFILIATIONS

Jordy Janssen orcid.org/0000-0003-1335-1312
Citizen Science Lab (Leiden University), NL; Research performed at Pulsaqua (WaterLab), NL
Marit Bogert orcid.org/0000-0002-0887-3565
TU Delft Science Centre (WaterLab), NL
Margaret Gold orcid.org/0000-0003-4853-2463
Citizen Science Lab (Leiden University), NL
Sandra de Vries orcid.org/0000-0002-7477-2447
Pulsaqua (WaterLab), NL

REFERENCES

Auerbach, J., Barthelmess, E.L., Cavalier, D., Cooper, C., Fenyk, H., Haklay, M, Hulbert, J.M., et al. (2019) The problem with delineating narrow criteria for citizen science, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(31), pp. 15336–15337. Available at: https://doi.org/10.1073/pnas.1909278116.

Becker-Klein, R., Peterman, K. and **Stylinski, C.** (2016) Embedded assessment as an essential method for understanding public engagement in citizen science, *Citizen Science: Theory and Practice*, 1(1), p. 8. Available at: https://doi.org/10.5334/cstp.15.

Bela, G., Peltola, T., Young, J.C., Balázs, B., Arpin, I., Pataki, G., Hauck, J., et al. (2018) The roles of scientists in participatory sustainability science: A case study from Hungary. *Conservation Biology*, 32(4), pp. 743–753. Available at: https://doi.org/10.1111/cobi.12762

Bieszczad, S.R., Fochler, M. and **Brodschneider, R.** (2023) How citizen scientists see their own role and expertise: An explorative study of the perspectives of beekeepers in a citizen science project, *Citizen Science: Theory and Practice*, 8(1). Available at: https://doi.org/10.5334/cstp.501.

Bogert, J., De Vries, S., Kunst, S., Nastase, N., Kloosterman, W., Knols, K., Schleiss, M., et al. (2022) Implementing citizen science within open science: Identifying extra-academic skills, collaborations, rewards and recognitions in the context of a university, *The Evolving Scholar* [Preprint]. Available at: https://doi.org/10.24404/62bad143c79c8753371e2366.

Bonney, R., Phillips, T.B., Ballard, H.L. and **Enck, J.W.** (2016) Can citizen science enhance public understanding of science? *Public Understanding of Science*, 25(1), pp. 2–16. Available at: https://doi.org/10.1177/0963662515607406

Brossard, D., Lewenstein, B., and **Bonney, R.** (2005) Scientific knowledge and attitude change: The impact of a citizen science project, *International Journal of Science Education*, 27(9), pp. 1099–1121. Available at: https://doi.org/10.1080/09500690500069483.

Bruyere, B. and **Rappe, S.** (2007) Identifying the motivations of environmental volunteers, *Journal of Environmental Planning and Management*, 50(4), pp. 503–516. Available at: https://doi.org/10.1080/09640560701402034.

Campbell, J.L., Quincy, C., Osserman, J., and **Pedersen, O.K.** (2013) Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement, *Sociological Methods and Research*, 42(3), pp. 294–320. Available at: https://doi.org/10.1177/0049124113500475.

Chase, S.K. and **Levine, A.** (2018) Citizen science: Exploring the potential of natural resource monitoring programs to influence environmental attitudes and behaviors, *Conservation Letters*, 11(2), pp. 1–10. Available at: https://doi.org/10.1111/conl.12382.

Constant, N. and **Roberts, L.** (2017) Narratives as a mode of research evaluation in citizen science: understanding broader science communication impacts, *Journal of Science Communication*, 16(04), p. A03. Available at: https://doi.org/10.22323/2.16040203.

Davis, C., Del Bianco, V., Peterman, K., Grover, A., Phillips, T. and **Becker-Klein, R.** (2022) Diverse and important ways evaluation can support and advance citizen science, *Citizen Science: Theory and Practice*, 7(1). Available at: https://doi.org/10.5334/cstp.482.

Feng, G.C. (2014) Intercoder reliability indices: Disuse, misuse, and abuse, *Quality and Quantity*, 48(3), pp. 1803–1815. Available at: https://doi.org/10.1007/s11135-013-9956-8.

Finger L., van den Bogaert V., Schmidt L., Fleischer J., Stadtler M., Sommer K., and **Wirth J.** (2023) The science of citizen science: A systematic literature review on educational and scientific outcomes, *Frontiers in Education*, 8(1226529). Frontiers Media SA. Available at: https://doi.org/10.3389/feduc.2023.1226529.

Gale, N.K., Heath, G., Cameron, E., Rashid, S., and **Redwood, S.** (2013) Using the framework method for the analysis of qualitative data in multi-disciplinary health research, *BMC Medical Research Methodology*, 13(1). Available at: https://doi.org/10.1186/1471-2288-13-117.

Gibbs, G. (2007) Thematic Coding and Categorizing, in: *Qualitative Research kit: Analyzing qualitative data*, London, England: SAGE Publications, Ltd. pp. 38–55. Available at: https://doi.org/10.4135/9781849208574.

Greving, H., Bruckermann, T., Schumann, A., Stillfried, M., Börner, K., Hagen, R., Kimmig, S.E., et al. (2023) Attitudes toward engagement in citizen science increase self-related, ecology-related, and motivation-related outcomes in an urban wildlife project, *BioScience*, 73(3), pp. 206–219. Available at: https://doi.org/10.1093/biosci/biad003.

Groulx, M., Brisbois, M.C., Lemieux, C.J., Winegardner, A., and **Fishback, L.** (2017) A role for nature-based citizen science in promoting individual and collective climate change action? A systematic review of learning outcomes, *Science Communication*, 39(1), pp. 45–76. Available at: https://doi.org/10.1177/1075547016688324.

Hackett, A. and **Strickland, K.** (2019) Using the framework approach to analyse qualitative data: a worked example, *Nurse researcher*, 26(2), pp. 8–13. Available at: https://doi.org/10.7748/nr.2018.e1580.

Haklay, M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., and **Vohland, K.** (2021) What is citizen science? The challenges of definition, in: Vohland, K. et al. *The Science of Citizen Science*. Springer International Publishing, pp. 13–33. Available at: https://doi.org/10.1007/978-3-030-58278-4_2.

Halpin, S.N. (2024) Inter-coder agreement in qualitative coding: Considerations for its use, *American Journal of Qualitative Research*, 8(3), pp. 23–43. Available at: https://doi.org/10.29333/ajqr/14887.

Haywood, B.K. (2014) A "sense of place" in public participation in scientific research, *Science Education*, 98(1), pp. 64–83. Available at: https://doi.org/10.1002/sce.21087.

Haywood, B.K. (2016) Beyond data points and research contributions: The personal meaning and value associated with public participation in scientific research, *International Journal of Science Education, Part B*, 6(3), pp. 239–262. Available at: https://doi.org/10.1080/21548455.2015.1043659.

Heigl, F., Kieslinger, B., Paul, K.T., Uhlik, J., and **Dörler, D.** (2019) 'Toward an international definition of citizen science', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, pp. 8089–8092. Available at: https://doi.org/10.1073/pnas.1903393116.

Jensen, E.A. and **Laurie, C.** (2016) 'Research Design', in *Doing Real Research: A Practical Guide to Social Research*. 1st edn. London: SAGE Publications Ltd, pp. 4–28.

Jordan, R.C., Ballard, H.L., and **Phillips, T.B.** (2012) Key issues and new approaches for evaluating citizen-science learning outcomes, *Frontiers in Ecology and the Environment*, 10(6), pp. 307–309. Available at: https://doi.org/10.1890/110280.

Jordan, R.C., Gray, S.A., Howe, D.V., Brooks, W.R., and **Ehrenfeld, J.G.** (2011) Knowledge gain and behavioral change in citizen-science programs, *Conservation Biology*, 25(6), pp. 1148–1154. Available at: https://doi.org/10.1111/j.1523-1739.2011.01745.x.

Jørgensen, F.A. and **Jørgensen, D.** (2021) Citizen science for environmental citizenship, *Conservation Biology*, 35(4), pp. 1344–1347. Available at: https://doi.org/10.1111/cobi.13649.

Kieslinger, B., Schäfer, T., Heigl, F., Dörler, D., Richter, A., and **Bonn, A.** (2018) Evaluating citizen science: Towards an open framework, in A. Bonn et al. (eds) *Citizen Science: Innovation in Open Science, Society and Policy*. UCL Press (Innovation in Open Science, Society and Policy), pp. 81–96. Available at: http://www.jstor.org/stable/j.ctv550cf2.13.

Kormos, C. and **Gifford, R.** (2014) The validity of self-report measures of proenvironmental behavior: A meta-analytic review, *Journal of Environmental Psychology*, 40, pp. 359–371. Available at: https://doi.org/10.1016/j.jenvp.2014.09.003.

Landis, J.R. and **Koch, G.G.** (1977) The measurement of observer agreement for categorical data, *Biometrics*, 33(1), p. 159. Available at: https://doi.org/10.2307/2529310.

Land-Zandstra, A., Agnello, G., and **Gültekin, Y.S.** (2021) Participants in citizen science, in K. Vohland et al. (eds) *The Science of Citizen Science*. Cham: Springer International Publishing, pp. 243–259. Available at: https://doi.org/10.1007/978-3-030-58278-4_13.

Land-Zandstra, A., Strick, M., Peeters, W., and **de Jong, P.** (2023) Het meten van impact van wetenschapscommunicatie: eindrapport IMPACTLAB. Available at: https://doi.org/10.5281/zenodo.7884383.

Linneberg, M.S. and **Korsgaard, S.** (2019) Coding qualitative data: a synthesis guiding the novice, *Qualitative Research Journal*, 19(3), pp. 259–270. Available at: https://doi.org/10.1108/QRJ-12-2018-0012.

Lynch, L.I., Dauer, J.M., Babchuk, W.A., Heng-Moss, T., and **Golick, D.** (2018) In their own words: The significance of participant perceptions in assessing entomology citizen science learning outcomes using a mixed methods approach, *Insects*, 9(1), pp. 1–15. Available at: https://doi.org/10.3390/insects9010016.

Marzi, G., Balzano, M., and **Marchiori, D.** (2024) K-Alpha Calculator–Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient, *MethodsX*, 12, p. 102545. Available at: https://doi.org/10.1016/j.mex.2023.102545.

McAteer, B., Flannery, W., and **Murtagh, B.** (2021) Linking the motivations and outcomes of volunteers to understand participation in marine community science, *Marine Policy*, 124(March 2020), p. 104375. Available at: https://doi.org/10.1016/j.marpol.2020.104375.

Mohajan, H.K. (2018) Qualitative research methodology in social sciences and related subjects, *Journal of Economic Development, Environment and People*, 7(1), pp. 23–48. Available at: https://doi.org/https://doi.org/10.26458/jedep.v7i1.571.

Örtengren, K. (2004) The Logical Framework Approach: A summary of the theory behind the LFA method, Sida Methods Development Unit. Available at: https://www.betterevaluation.org/tools-resources/summary-theory-behind-lfa-method.

Pateman, R.M. and **West, S.E.** (2023) citizen science: Pathways to impact and why participant diversity matters, *Citizen Science: Theory and Practice*, 8(1). Available at: https://doi.org/10.5334/cstp.569.

Peter, M., Diekötter, T., Höffler, T., and **Kremer, K.** (2021) Biodiversity citizen science: Outcomes for the participating citizens, *People and Nature*, 3(2), pp. 294–311. Available at: https://doi.org/10.1002/pan3.10193.

Peter, M., Diekötter, T. and **Kremer, K.** (2019) Participant outcomes of biodiversity citizen science projects: A systematic literature review, *Sustainability (Switzerland)*, 11(10), pp. 1–18. Available at: https://doi.org/10.3390/su11102780.

Phillips, T., Faulkner, H., Ferguson, M., Minarcheck, M., Porticella, N., and **Bonney, R.** (2017) Interactive user's guide for evaluating learning outcomes in citizen science. Cornell Lab of Ornithology, Ithaca, NY. Available at: www.citizenscience.org.

Phillips, T., Ferguson, M., Minarchek, M., Porticella, N., and **Bonney, R.** (2014) User's guide for evaluating learning outcomes from citizen science. Cornell Lab of Ornithology, Ithaca, NY. Available at: https://participatorysciences.org/resources/research-evaluation/ (Accessed: 3 April 2024).

Phillips, T.B., Ballard, H.L., Lewenstein, B.V., and **Bonney, R.** (2019) Engagement in science through citizen science: Moving beyond data collection, *Science Education*, 103(3), pp. 665–690. Available at: https://doi.org/10.1002/sce.21501.

Phillips, T.B., Porticella, N., and **Bonney, R.E.** (2018) A framework for articulating and measuring individual learning outcomes from participation in citizen science, *Citizen Science: Theory and Practice*, 3(2), p. 3. Available at: https://doi.org/10.5334/cstp.126.

Robinson, L.D., Cawthray, J.L., West, S.E., Bonn, A. and **Ansine, J.** (2018) Ten principles of citizen science, in A. Bonn, S. Hecker, M. Haklay, A. Bowser, Z. Makuch, and J. Vogel (eds.), *Citizen Science: Innovation in Open Science, Society and Policy* (pp. 27–40). UCL Press. Available at: http://www.jstor.org/stable/j.ctv550cf2.9

Schäfer, T., Kieslinger, B., Brandt, M., and **van den Bogaert, V.** (2021) 'Evaluation in Citizen Science: The Art of Tracing a Moving Target', in K. Vohland et al. (eds) *The Science of Citizen Science*. Cham: Springer International Publishing, pp. 495–514. Available at: https://doi.org/10.1007/978-3-030-58278-4_25.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., and **Simons, A.** (1991) Ease of retrieval as information: Another look at the availability heuristic, *Journal of Personality and Social Psychology*, 61(2), pp. 195–202. Available at: https://doi.org/10.1037/0022-3514.61.2.195.

Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., et al. (2012) Public participation in scientific research: A framework for deliberate design, *Ecology and Society*, 17(2). Available at: https://doi.org/10.5751/ES-04705-170229.

Shirk, J.L. and **Bonney, R.** (2018) Scientific impacts and innovations of citizen science, in S. Hecker et al. (eds) *Citizen Science*. UCL Press (Innovation in Open Science, Society and Policy), pp. 41–51. Available at: http://www.jstor.org/stable/j.ctv550cf2.10.

Sickler, J., Messick Cherry, T., Allee, L., Rice Smyth, R., and **Losey, J.** (2014) Scientific value and educational goals: Balancing priorities and increasing adult engagement in a citizen science project, *Applied Environmental Education and Communication*, 13(2), pp. 109–119. Available at: https://doi.org/10.1080/1533015X.2014.947051.

Somerwill, L. and **Wehn, U.** (2022) How to measure the impact of citizen science on environmental attitudes, behaviour and knowledge? A review of state-of-the-art approaches, *Environmental Sciences Europe*, 34(1), pp. 1–13. Available at: https://doi.org/10.1186/s12302-022-00596-1.

Syed, M. and **Nelson, S.C.** (2015) Guidelines for establishing reliability when coding narrative data, *Emerging Adulthood*, 3(6), pp. 375–387. Available at: https://doi.org/10.1177/2167696815587648.

Turaga, R.M.R., Howarth, R.B., and **Borsuk, M.E.** (2010) Pro-environmental behavior: Rational choice meets moral motivation, *Annals of the New York Academy of Sciences*, 1185, pp. 211–224. Available at: https://doi.org/10.1111/j.1749-6632.2009.05163.x.

Tversky, A. and **Kahneman, D.** (1973) Availability: A heuristic for judging frequency and probability, *Cognitive Psychology*, 5(2), pp. 207–232. Available at: https://doi.org/10.1016/0010-0285(73)90033-9.

von Gönner, J., Herrmann, T.M., Bruckermann, T., Eichinger, M., Hecker, S., Klan, F., Lorke, J., et al. (2023) Citizen science's transformative impact on science, citizen empowerment and socio-political processes, *Socio-Ecological Practice Research*, 5(1), pp. 11–33. Available at: https://doi.org/10.1007/s42532-022-00136-4.

WaterLab (2024) *Delft Meet Regen*. Available at: https://www.tudelft.nl/en/scd/waterlab/join-our-research/project-7-delft-measures (Accessed: 15 April 2024).

Wehn, U., Gharesifard, M., Ceccaroni, L., Joyce, H., Ajates, R., Woods, S., Bilbao, A., et al. (2021) 'Impact assessment of citizen science: state of the art and guiding principles for a consolidated approach', *Sustainability Science*, 16(5), pp. 1683–1699. Available at: https://doi.org/10.1007/s11625-021-00959-2.

Zaiontz, C. (no date) *Cohen's Kappa*, *real statistics using Excel*. Available at: https://real-statistics.com/reliability/interrater-reliability/cohens-kappa/ (Accessed: 13 November 2024).

TO CITE THIS ARTICLE:
Janssen, J., Bogert, M., Gold, M. and de Vries, S. 2025. Evaluating Participant Outcomes in the Weather Monitoring Citizen Science Project "Delft Meet Regen". *Citizen Science: Theory and Practice*, 10(1): 20, pp. 1–14. DOI: https://doi.org/10.5334/cstp.757

Submitted: 16 April 2024 **Accepted:** 09 May 2025 **Published:** 30 June 2025