Citizen science projects require agreement: a Delphi study to identify which knowledge on urban ecology is considered relevant from scientists’ and citizens’ perspectives

Till Bruckermann a,b, Milena Stillfriedc, Tanja M. Strakad and Ute Harms b

aInstitute of Education, Leibniz University Hannover, Hannover; bIPN Leibniz Institute for Science and Mathematics Education, Kiel; cLeibniz Institute for Zoo and Wildlife Research, Berlin; dDepartment of Ecology, Technische Universität Berlin, Berlin

ABSTRACT
In citizen science (CS) projects, acquired knowledge about a specific topic is the most frequently acknowledged learning outcome. However, whether both citizens and scientists perceive the same knowledge to be relevant to citizens’ learning in such projects remains unknown. Thus, establishing coherence between citizens’ information needs and scientists’ intentions to inform as well as exploring the reasons for why some knowledge is more relevant from citizens’ and scientists’ perspectives could help to achieve agreement regarding what knowledge is relevant for learning outcomes on the side of the citizens. By using the Delphi technique, we accounted for scientists’ and citizens’ perspectives on the relevance of knowledge in three fields of research on urban ecology. In our Delphi study, an emerging consensus indicated an overlap in relevance among the experts. We then analyzed two dimensions of relevance, that is, to whom and for what the knowledge is relevant. Our analyses of the dimensions revealed that consensus was more likely when we accounted for content-related differences and structural differences such as the communicatory perspective. When we accounted for content-related differences, relevance was higher for problem-oriented knowledge; therefore, this should be the focus of CS projects that are designed to achieve learning outcomes.

CONTACT Till Bruckermann till.bruckermann@iew.uni-hannover.de Institute of Education, Leibniz University Hannover, Hannover; IPN Leibniz Institute for Science and Mathematics Education, Kiel

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
whom the knowledge is relevant (i.e. scientists, citizens) and for what the knowledge is relevant (e.g. everyday life, academic research; Mayoh & Knutton, 1997). Previous research indicated that projects that aim to promote citizens’ knowledge should account not only for knowledge that is relevant from scientists’ perspectives but also for knowledge that citizens and scientists find equally relevant (Stocklmayer & Bryant, 2012). Hence, further research is required that more precisely explores the conditions under which knowledge about the project’s research topic is considered relevant from both scientists’ and citizens’ perspectives. Information about these conditions could then be used to specifically design CS projects to promote citizens’ learning. The current study investigated which knowledge about the research topic of three urban ecology CS projects—concerning bats, terrestrial mammals, and air pollution—was relevant from citizens’ and scientists’ perspectives, and it examined the conditions under which it was more likely that the perceptions of relevant knowledge would agree.

**Theoretical background**

In the recent past, citizens’ individual learning outcomes in CS projects have become increasingly important (Bonney et al., 2016; Jordan et al., 2012; Wals et al., 2014). Most CS projects focus on the promotion of citizens’ content knowledge about the project’s research topic as an individual learning outcome (i.e. topic-specific knowledge; Groulx et al., 2017; Peter et al., 2019) because those projects assume that enhanced knowledge leads to more positive attitudes toward science (Brossard et al., 2005). This knowledge, however, does not affect citizens’ attitudes toward science because it is seemingly less relevant to them (Stocklmayer & Bryant, 2012). Therefore, we explored two dimensions of relevance in order to establish which knowledge is commonly perceived as being relevant from both citizens’ and scientists’ perspectives. Questions asked about the specific topic served as indicators for relevant knowledge and a Delphi technique established which knowledge was considered to be relevant from both citizens’ and scientists’ perspectives.

**Promoting topic-specific knowledge in citizen science projects**

Topic-specific knowledge refers to knowledge about the content of a specific research topic in CS projects (Aristeidou & Herodotou, 2020). Literature reviews indicate that topic-specific knowledge is among the most common individual learning outcome of CS projects, for example, topic-specific knowledge about climate change (Groulx et al., 2017), about the environment (Phillips et al., 2018; Schuttler et al., 2018), about biodiversity (Peter et al., 2019), and about different topics in online CS projects (Aristeidou & Herodotou, 2020). The reason that so many CS projects focus on the promotion of knowledge is that enhanced knowledge is assumed to promote more positive attitudes toward science (Brossard et al., 2005). Yet, in general, research findings on the relationship between knowledge and attitudes toward science are mixed (see Allum et al., 2008, for an overview).

In CS projects, findings on the relationship between knowledge and attitudes toward science are also mixed (Brossard et al., 2005; Jordan et al., 2011; cf. Bruckermann et al., 2021). One reason for this could be that CS projects promote knowledge that is relevant to academic scientists but not necessarily to the participating citizens’ everyday life (Haywood & Besley, 2014; Lewenstein & Bonney, 2004). Other research suggests that ‘more “local” types of knowledge’ strengthen the relationship with attitudes (Allum et al., 2008, p. 51) because they are embedded in the local context, for example, of a community CS project (Haywood, 2016). Currently, CS projects promote knowledge that is seemingly less relevant to citizens (Haywood & Besley, 2014). For informal learning, such as in CS projects, research should, therefore, explore which knowledge is relevant not only to academic scientists but also to citizens’ everyday life (Stocklmayer et al., 2010; Stocklmayer & Bryant, 2012). Relevance has at least two dimensions, which concern the question of (1) to whom the knowledge is relevant (i.e. scientists, citizens) and (2) for what the knowledge is relevant (e.g. everyday life,
academic research; Mayoh & Knutton, 1997; see Aikenhead, 2003; Stuckey et al., 2013, for an overview).

**Two dimensions of relevance of knowledge**

The first dimension of relevance concerns the question of to whom knowledge is relevant (Mayoh & Knutton, 1997), for example, whether scientists or citizens deem the respective knowledge to be relevant. This dimension is related to structural differences in the relevance of knowledge. For a long time, communication with citizens about scientific knowledge was limited to a deficit model with a unidirectional approach that considered scientists as senders of information and citizens as recipients who lacked information (Trench, 2008). Within this unidirectional approach, scientists provided citizens with knowledge that the scientists themselves considered relevant or, at least, with knowledge that, from their point of view, could be relevant to citizens (Carr et al., 2017). When using the deficit model in informal learning, scientific knowledge that is relevant to academic scientists is prioritized over knowledge that is relevant to citizens (Haywood & Besley, 2014; Lewenstein & Bonney, 2004). Furthermore, academic scientists’ capabilities in estimating citizens’ prerequisites are limited (Besley et al., 2015; Carr et al., 2017; cf. Bromme et al., 2001), especially when it comes to resources for project design in CS (Bonney et al., 2016). More recently, dialogic and participatory models of science communication that consider knowledge that is relevant to both citizens and scientists have been used, as these models are more audience-centric (Besley et al., 2015; Carr et al., 2017). In the current study, we explored structural differences such as the sender-recipient perspective when investigating the relevance of knowledge because dialogic and participatory approaches account for knowledge that is relevant to both citizens and scientists.

The second dimension of relevance concerns the question of what knowledge is relevant for (Mayoh & Knutton, 1997), for example, for everyday life or for academic research. This dimension concerns the content-related differences in the relevance of knowledge. While knowledge that is relevant for educational certification is canonically organized, such as the knowledge that is learned in formal contexts, knowledge that is relevant for everyday life appears to be more problem-oriented, such as the knowledge that is learned in informal contexts (Aikenhead, 2003; Bromme & Kienhues, 2017). Moreover, canonical knowledge cannot be directly translated into use for everyday situations (Aikenhead, 2003). In informal learning contexts, citizens seek information that will help them to solve their problems (Besley, 2014; Hendriks et al., 2016). Previous research showed that exploring specific issues or problems extended and deepened knowledge that is relevant to citizens who engaged with science (Bromme & Goldman, 2014; Ryder, 2001). These issues often included non-scientific aspects such as political, ethical, or religious considerations. Therefore, citizens’ search for information is mixed with normative goals so that the acquired knowledge fits the individual’s beliefs and values (de Groot et al., 2013). In the current study, we thus explored content-related differences, such as the problem-orientation of topics, when investigating the relevance of knowledge because relevance seems to depend on the respective situation in which the knowledge is applied.

**Identifying common perceptions of relevant knowledge in CS projects**

Questions that citizens ask about a topic are indicators of topic-specific knowledge that is relevant to the questioners. Questions from museum visitors, internet users, students, and other individuals are gaining attention in educational research because they are indicative of individuals’ attempts to gain knowledge about a topic (e.g. in the museum: Seakins, 2015; on the internet: Baram-Tsabari et al., 2009). Moreover, questions indicate which precise topics are relevant to the interests of the individual questioners (Baram-Tsabari et al., 2006). Informal learning could take up the interests of citizens by accounting for the questions that individuals ask (Seakins, 2015). To do this, research systematically explores questions asked by citizens to uncover patterns of topics that citizens want to know about. Questions have already guided the development of curricula in formal
learning contexts and the co-curation of exhibitions in informal learning contexts (Seakins, 2015). Similarly, the participatory approach of CS projects has the potential to align informal learning with the questions asked by citizens (Bonney et al., 2016) and, thus, to move from a unidirectional deficit model towards a bidirectional dialogic or multidirectional participatory model (Carr et al., 2017; Trench, 2008). For CS projects to explore questions that are relevant to citizens’ lives, the citizens’ questions need to be included in the decision-making process concerning the projects’ objectives (Bonney et al., 2016).

The Delphi technique helps to identify common perceptions of what is considered to be relevant knowledge by accounting for questions on the topic from different perspectives. The different perspectives on the questions could be explored by the Delphi technique to determine structural differences, that is, to whom knowledge is relevant, and content-related differences, that is, for what knowledge is relevant. The Delphi technique is suitable to account for different dimensions of relevance (Aikenhead, 2003), that is, the multidimensionality of relevance (Stuckey et al., 2013). It facilitates the decision-making process among different expert groups such as academic scientists and societal experts (e.g. Blanco-López et al., 2015). The Delphi technique uses the expertise of experts to make decisions on critical questions in fields such as education (e.g. Clayton, 1997) as well as ecology and conservation (e.g. Mukherjee et al., 2015). The Delphi technique can be used for decisions as well as in arguments (Mukherjee et al., 2015). While the decision Delphi is a technique that is used to establish a common ground of opinions, the argument Delphi is a technique that is used to reveal reasons for divergent opinions (Mukherjee et al., 2015). Hence, when the Delphi technique is used to decide on program goals, it is called a decision Delphi. Previous research showed that, for example, program goals in a natural history museum could be adapted to the perspectives of researchers and practitioners by using the Delphi technique (Seakins & Dillon, 2013). Therefore, the current study explored which questions about topics on urban ecology in CS projects are relevant from both scientists’ and citizens’ perspectives by applying the Delphi technique.

The current study involved academic scientists and societal experts in an iterative and collective approach by using the Delphi technique to gain information on which topic-specific knowledge in CS projects—whose aim is to promote individual learning outcomes—is relevant to both citizens and scientists. This approach focused on questions as indicators of relevant knowledge and explored why some questions were commonly perceived as relevant whereas others were not. More precisely, the current study examined the extent to which questions were relevant to both scientists and citizens when two dimensions of relevance were considered, that is,

1. To whom are the questions relevant (i.e. structural differences)? and
2. For what are the questions relevant (i.e. content-related differences)?

Thus, the twofold aim of this study was (1) to reach a consensus on knowledge that is relevant to both citizens and scientists, and (2) to explore the extent to which this consensus depends on structural and content-related differences. Using the Delphi technique, we first explored the questions for which there can be a consensus on the relevance of knowledge from scientists’ and citizens’ perspectives (i.e. decision Delphi), and, secondly, we explored reasons for divergent perceptions of relevance depending on structural and content-related differences (i.e. argument Delphi; Mukherjee et al., 2015). The Delphi technique made it possible to account for different perspectives by relying on experts from two professional groups, that is, academic scientists and societal experts who have extensive contact with the public.

**Methods**

The current study applied the Delphi technique, first, to establish a consensus on specific themes (i.e. decision Delphi) and, second, to explore structural and content-related differences that probably lead to divergent perceptions of the relevance of knowledge (i.e. argument Delphi). In the
iterative process of the Delphi technique, we intended three rounds for each of the three topics (i.e. bats, terrestrial mammals, and air pollution). For the analysis of experts’ answers to the questionnaires of the Delphi, we used qualitative content analysis (QCA) to paraphrase, reduce, and categorize experts’ statements (see Qualitative content analysis section) as well as descriptive and inferential analyses to explore differences in the experts’ ratings (see Statistical analysis section).

**Delphi technique**

The reliability and validity of the Delphi technique strongly depend on the adherence to methodological criteria such as selecting participants based on their expertise, defining consensus criteria in advance, formulating stopping criteria, and creating rules for item dropping (Clayton, 1997; Diamond et al., 2014). Adhering to and reporting quality criteria enhances the traceability and reproducibility for other researchers. The flow chart in Figure 1 reflects the three Delphi rounds for each topic, and the following sections document the methodological criteria for the Delphi, such as the selection of an expert panel before the first round.

**Panels of expert participants**

Surveys that use the Delphi technique differ from other survey types due to the expertise within the sample. People in the sample are experts with knowledge and experience in the respective topic (Clayton, 1997). The expert sample for our study was composed of academic scientists and societal experts so that both scientists’ and citizens’ perspectives were represented. Hence, the expert panels were heterogeneous in their professional stratification (i.e. university academics, nature conservationists from non-governmental organizations [NGO], and civil servants from state authorities). The criteria for selecting experts who we label as academic scientists included that they had to specialize in the respective field (e.g. involvement in the topic bat ecology), currently be actively pursuing a research position in this field, and belong to a university or research institution. The criteria for selecting experts who we label as societal experts included that they were affiliated

---

**Figure 1.** Flow Chart for the Delphi Technique Indicating the Total Number of Experts Involved \((N)\) and the Number of Questions for Analysis \((n)\). *Note.* QCA = qualitative content analysis.
with NGOs or state authorities and within responsible for the topic of interest (e.g. for the topic of terrestrial mammals: state foresters), and that they frequently communicated with citizens regarding the topic; thus, their professional status made them experts on citizens’ different views and questions regarding the topics at hand (e.g. Osborne et al., 2003; Seakins & Dillon, 2013). Furthermore, all experts were not allowed to be associated with the CS project studied here, so we can rule out any bias in the responses. The panel size was established to account for the heterogeneity as well as constraints such as response rate (i.e. five to 10 people suffice in heterogeneous samples; Clayton, 1997). Three expert panels were sought (one expert panel per topic; \( N = 33 \); see Figure 1). All experts were invited to participate by scientists from the research teams of the present CS project and they were informed about the objectives of the Delphi study. They agreed to become part of one of the expert panels before being contacted by email with the Delphi questionnaires.

**Delphi questionnaires**

The experts were provided with a questionnaire, which was further developed based on the results of the previous round. The questionnaire in the first round asked the experts to name up to 10 questions on the respective topic (i.e. urban ecology in cities concerning bats, terrestrial mammals, and air pollution) that are relevant to citizens because the questions mirror what individuals would like to know (Baram-Tsabari et al., 2009). Furthermore, to identify common questions for a dialogic approach, experts were asked to indicate whether scientists want citizens to know about the question (i.e. sender perspective) or whether citizens themselves want to know about it (i.e. recipient perspective; Carr et al., 2017). The questionnaires of the second and third rounds then asked the experts to rate how relevant each question is in order for citizens to know more about the topic. In the questionnaires, experts rated the relevance of questions on a 5-point Likert scale ranging from 1 (not relevant) to 5 (relevant). All questions without consensus in the second round were included in the questionnaire of the third round for another rating of relevance. For each question, the median of all experts’ ratings and the respective individual expert’s rating from the second round was indicated.

**Consensus definition**

We defined consensus as the agreement of academic scientists and societal experts concerning the relevance of a question from the perspectives of both citizens and scientists. As it represents a standard threshold of consensus (Diamond et al., 2014), a percent agreement in which more than half of the expert panel (> 50%) gave the same rating served as the cutoff for consensus.

**Stopping criterion**

Although consensus on which questions were relevant for all three topics was desirable, non-consensual questions could lead to conclusions with different points of view. Therefore, the number of rounds was limited to three to prevent participant dropout (Clayton, 1997).

**Question dropping**

Questions were dropped or merged after the first round based on inductive categorization in a qualitative content analysis. We dropped or merged questions into a single question if they met one of the following two criteria: the question was mentioned at least twice by different experts (duplicates) or another question was more specific than the original. After the second round, questions were dropped if consensus had already been established. By reducing the number of questions that the experts had to assess, we aimed to reduce the workload for the experts and to strengthen the willingness to participate in a further round. The reduction of questions from round to round reduces expert dropout.
Qualitative content analysis

All questions collected in the first round were grouped into categories based on an inductive approach of a qualitative content analysis (QCA; Mayring, 2000). Three independent research teams that were recruited from scientists in the present CS project categorized the questions on the respective topics (bats: four raters; terrestrial mammals: four raters; air pollution: three raters). They were instructed by a manual on the inductive approach of QCA and trained in a face-to-face meeting with the first author. The teams discussed the categories for the formative assessment of reliable category application after categorizing half of the questions. For the final categorization of all questions, a catalog of questions with a category description and anchor examples was established. After the raters of the respective teams independently categorized all of the questions, we calculated Krippendorff’s α to establish the summative assessment of interrater reliability (Krippendorff, 2004). The interpretation of Krippendorff’s α revealed substantial agreement in the categorization of questions (Krippendorff’s α = 0.73; 0.77; see Landis & Koch, 1977, for comparable thresholds).

Statistical analysis

In addition to establishing a consensus by using the technique of a decision Delphi, we further analyzed the questions on the three topics (bats [BT]: \( n_{BT} = 65 \); terrestrial mammals [TM]: \( n_{TM} = 61 \); air pollution [AP]: \( n_{AP} = 69 \)) by conducting variance analyses (ANOVA) to establish quantitative differences in the relevance ratings across factors such as authors, perspective, consensus, and categories (i.e. revealing reasons for different opinions by using the technique of an argument Delphi). We used paired t-tests to compare the means of relevance ratings, that is, the mean relevance for all questions on the three topics (\( n_{BT} = 65 \); \( n_{TM} = 61 \); \( n_{AP} = 69 \)) between both expert groups (academic scientists and societal experts). We statistically tested quantitative differences in the means of relevance ratings (i.e. 5-point Likert scale) by employing an exploratory approach to indicate structural and content-related differences but not to test any hypotheses.

Results

In the first round, a total of 118 questions on the topic of BT, 99 questions on the topic of TM, and 108 questions on the topic of AP were collected. Three independent research teams from the respective fields analyzed the questions on each topic and merged similar questions (see Qualitative content analysis section above): \( n_{BT} = 65 \) questions on BT, \( n_{TM} = 61 \) questions on TM, and \( n_{AP} = 69 \) questions on AP were retained for the second round (see Figure 1, for a flow chart detailing the numbers of questions).

Emerging themes from qualitative content analysis after first round

The inductive qualitative analysis of the questions from the first round by the three independent research teams revealed 11 distinct categories for the topic of BT, 8 categories for the topic of TM, and 11 categories for the topic of AP. The numbers after the acronyms refer to the respective categories of the three topics, but they do not reflect any ranking (e.g. BT1 refers to the first category on the topic of BT). Clear-cut definitions with anchoring examples distinguished the categories that followed from an analytical point of view according to the methodology of QCA (Mayring, 2000; see Appendix A: Tables 1–3). However, from a holistic point of view, overarching themes emerged from the categories of the three topics (e.g. health effects: BT3: threats; TM3: diseases [zoonosis]; AP4: health effects); more specific themes emerged that were only shared between two of the topics (e.g. protection: BT6: protection and mitigation strategies; TM6: protection of biodiversity); and specific themes emerged for each topic (e.g. BT9: systematics and morphology). In the text, Tables 1–3 provide an overview of the categories for each topic (i.e. BT, TM, and AP, respectively; see
Appendix A: Tables 1–3, for category definitions and exemplary questions). The columns in the tables provide the number \((n)\) of questions per category for each of the three Delphi rounds. The first two columns (under first round) indicate how many questions the experts initially suggested \((n)\), and how many questions were merged within the different categories based on the qualitative content analysis later on \((n_{\text{merged}})\). The columns for the second and third Delphi rounds include the means \((M)\) and standard deviations \((SD)\) of the relevance ratings for each category. Categories whose questions were more relevant on average are higher up in the tables.

**Overarching themes**

Experts proposed questions that were related to threats to human health, legal regulations, protection strategies, and spatial distribution patterns in the city for all three topics. In the case of bats (i.e. category BT3) and of terrestrial mammals (i.e. category TM3), questions covered threats of diseases (zoonosis) transmitted by animals to humans. In the case of air pollution (i.e. categories AP4 and AP11), questions covered the different health effects of air pollution and toxic substances such as black carbon and nitric oxides. The following example illustrates the thematic overlap of questions:

‘How great is the risk of being infected with rabies by bats?’ (BT3); ‘Is there any risk of becoming infected with diseases from wild animals when eating vegetables from the garden?’ (TM3); ‘What is the correlation between air pollution and public health?’ (AP4).

Furthermore, questions covered legal regulations to protect bats (i.e. category BT7), manage wildlife (i.e. category TM8), or reduce air pollution (i.e. category AP1), and measures that one can take to protect wild animals in cities (i.e. categories BT6 and TM6) or to protect oneself from air pollution (i.e. categories AP2 and AP7).

Lastly, one overarching theme related to the spatial distribution patterns of wild animals (habitats: categories BT4 and TM4) or air pollution in cities (spatiotemporal variability and sources of air pollution: categories AP8 and AP9). These questions addressed how animals find places to live in the city or how air pollution is distributed within the city.

**Specific themes**

Inductive categorization with a qualitative content analysis revealed specific themes that were apparent in only one of the three topics of bats, terrestrial mammals, and air pollution. While

### Table 1. Categories for Bats (BT) and Their Relevance Based on Means \((M)\) and Standard Deviation \((SD)\) for Each Round.

| Category                          | first round | second round | third round |
|----------------------------------|-------------|--------------|-------------|
|                                  | n | \(n_{\text{merged}}\) | \(N\) | \(M\) | \(SD\) | \(n\) | \(M\) | \(SD\) | \(n\) | \(M\) | \(SD\) | \(n\) | \(M\) | \(SD\) |
| BT3: threats                     | 15 | 10 | 5 | 5.0 | 0.00 | 5 | 5.0 | 0.00 | 2 | 2 | 2 | 2 | 2 | 2 |
| BT5: damages                     | 4 | 2 | 2 | 5.0 | 0.00 | 2 | 5.0 | 0.00 | 2 | 2 | 2 | 2 | 2 | 2 |
| BT7: legal basis of protection   | 15 | 5 | 10 | 4.7 | 0.54 | 7 | 5.0 | 0.00 | 3 | 4.3 | 0.58 | 1 | 5.0 | 0.00 |
| BT6: protection and mitigation strategies | 13 | 6 | 7 | 4.6 | 0.75 | 5 | 5.0 | 0.00 | 2 | 4.0 | 1.41 | 1 | 5.0 | 0.00 |
| BT10: handling encounter situations | 16 | 9 | 7 | 4.6 | 0.75 | 5 | 5.0 | 0.00 | 2 | 3.8 | 0.35 | 0 | — | — |
| BT2: information                 | 9 | 4 | 5 | 4.5 | 0.87 | 3 | 5.0 | 0.00 | 2 | 4.0 | 1.41 | 2 | 4.0 | 1.41 |
| BT8: ecosystem services          | 7 | 4 | 3 | 4.5 | 0.87 | 2 | 5.0 | 0.00 | 1 | 4.0 | — | 0 | — | — |
| BT4: habitat                     | 16 | 5 | 11 | 4.0 | 0.89 | 2 | 5.0 | 0.00 | 9 | 3.5 | 1.12 | 6 | 3.7 | 1.21 |
| BT1: ecology                     | 15 | 5 | 10 | 3.5 | 0.69 | 0 | — | — | 10 | 3.5 | 0.71 | 3 | 3.0 | 1.00 |
| BT9: systematics and morphology  | 6 | 2 | 4 | 2.5 | 0.58 | 2 | 2.5 | 0.71 | 2 | 2.5 | 0.71 | 1 | 2.0 | — |
| BT11: mythology                  | 2 | 1 | 1 | 2.5 | — | 0 | — | — | 1 | 3.0 | — | 0 | — | — |
| Sum                              | 118 | 53 | 65 | 4.2 | 0.95 | 33 | 4.8 | 0.62 | 32 | 3.6 | 0.90 | 14 | 3.6 | 1.22 |

Note. The order of categories is based on the mean from the second round; 1single case, no variance; 2consensus already reached in second round.
some questions dealt with ecosystem services and the mythology of bats (i.e. categories BT8 and BT11), questions on terrestrial mammals covered animal nutrition in cities (TM1). Specific themes in questions on air pollution focused on the measurement and distinctive signs of air pollution (i.e. categories AP5 and AP3).

Analysis of consensus on relevance in the second and third rounds

Consensus was established in the second round for \( n = 33 \) questions concerning bats (of \( n_{BT} = 65 \); see Table 1), \( n = 30 \) questions concerning terrestrial mammals (of \( n_{TM} = 61 \); see Table 2), and \( n = 32 \) concerning air pollution (of \( n_{AP} = 69 \); see Table 3). The third round established consensus on another \( n_{BT} = 14 \), \( n_{TM} = 19 \), and \( n_{AP} = 19 \) questions. At the end of the Delphi survey, there was still no consensus for a certain number of questions. There was dissent on \( n_{BT} = 18 \), \( n_{TM} = 12 \), and \( n_{AP} = 18 \) questions.

To obtain data relevant to our initial question of why some questions were more commonly perceived as being relevant than others, we exploratorily analyzed structural differences that were related to (a) how different expert groups perceived relevance (raters: academic scientists vs. societal experts), (b) whose questions were perceived as being relevant (authors: those of academic scientists

Table 2. Categories for Terrestrial Mammals (TM) and Their Relevance Based on Means (\( M \)) and Standard Deviation (\( SD \)) for Each Round.

|                  | first round | second round | third round |
|------------------|-------------|--------------|-------------|
|                  | \( n \)     | \( n_{merg} \) | \( n \)     | \( M \) | \( SD \) | \( n \)     | \( M \) | \( SD \) | \( n \)     | \( M \) | \( SD \) |
| TM2: help desks  | 3           | 0            | 3           | 5.0   | 0.00  | 3           | 5.0   | 0.00  | 0          | 0.00  | 0.00  |
| TM5: damages/damage prevention | 14 | 4 | 10 | 4.7 | 0.42 | 8 | 4.8 | 0.46 | 2 | 4.5 | 0.71 |
| TM7: handling encounter situations | 25 | 15 | 10 | 4.7 | 0.47 | 7 | 4.9 | 0.38 | 3 | 4.3 | 0.58 |
| TM3: diseases (zoonosis) | 13 | 6 | 7 | 4.5 | 0.50 | 3 | 5.0 | 0.00 | 4 | 4.3 | 0.50 |
| TM8: wildlife management | 7 | 1 | 6 | 4.2 | 0.41 | 2 | 4.5 | 0.71 | 4 | 4.0 | 0.00 |
| TM6: protection of biodiversity | 13 | 5 | 8 | 4.1 | 0.64 | 1 | 5.0 | — 1 | 7 | 4.1 | 0.90 |
| TM1: nutrition | 10 | 4 | 6 | 3.9 | 1.50 | 3 | 3.7 | 2.31 | 3 | 4.0 | 0.00 |
| TM4: urban habitats | 14 | 3 | 11 | 3.7 | 0.46 | 3 | 3.7 | 0.58 | 8 | 3.8 | 0.53 |
| Sum              | 99          | 38           | 61          | 4.3   | 0.73 | 30 | 4.6 | 0.86 | 31 | 4.1 | 0.58 |

Note. The order of categories is based on the mean from the second round; 1single case, no variance; 2consensus already reached in second round.

Table 3. Categories for Air Pollution (AP) and Their Relevance Based on Means (\( M \)) and Standard Deviation (\( SD \)) for Each Round.

|                  | first round | second round | third round |
|------------------|-------------|--------------|-------------|
|                  | \( n \)     | \( n_{merg} \) | \( n \)     | \( M \) | \( SD \) | \( n \)     | \( M \) | \( SD \) | \( n \)     | \( M \) | \( SD \) |
| AP8: sources of AP | 9           | 0            | 3           | 5.0   | 0.00  | 3           | 5.0   | 0.00  | 2          | 4.5 | 0.71 |
| AP2: alternative mobility | 4 | 0 | 3 | 5.0 | 0.50 | 3 | 5.0 | 0.00 | 3 | 4.3 | 0.58 |
| AP7: protection against AP | 19 | 11 | 8 | 4.6 | 0.52 | 6 | 4.8 | 0.41 | 2 | 4.0 | 0.00 |
| AP11: toxic substances in air | 6 | 2 | 4 | 4.3 | 0.96 | 2 | 5.0 | 0.00 | 2 | 3.5 | 0.71 |
| AP4: health effects | 13 | 2 | 11 | 4.2 | 1.25 | 7 | 4.4 | 1.51 | 4 | 3.8 | 0.50 |
| AP6: meteorology | 13 | 6 | 7 | 4.1 | 0.69 | 2 | 5.0 | 0.00 | 5 | 3.8 | 0.45 |
| AP1: air quality regulations | 13 | 4 | 9 | 4.0 | 0.87 | 3 | 5.0 | 0.00 | 6 | 3.5 | 0.55 |
| AP5: measurements | 10 | 2 | 8 | 3.9 | 0.99 | 3 | 4.7 | 0.58 | 5 | 3.5 | 0.51 |
| AP10: species of AP | 8 | 3 | 5 | 3.8 | 1.10 | 2 | 5.0 | 0.00 | 3 | 3.0 | 0.00 |
| AP3: distinctive signs of AP | 5 | 2 | 3 | 3.7 | 0.58 | 1 | 4.0 | — 1 | 2 | 3.5 | 0.71 |
| AP9: spatiotemporal variability of AP | 8 | 1 | 7 | 3.6 | 0.54 | — | — | — | 7 | 3.6 | 0.53 |
| Sum              | 108         | 39           | 69          | 4.1   | 0.89 | 32 | 4.8 | 0.75 | 37 | 3.6 | 0.52 |

Note. The order of categories is based on the mean from the second round; 1single case, no variance; 2consensus already reached in second round.
vs. societal experts vs. both), (c) how relevance differed between perspectives (relevant to the perspective: of a sender vs. a recipient vs. both), and (d) whether consensus was established (consensus: second vs. third round vs. none). We also analyzed content-related differences, that is, (e) whether the overarching themes of the questions differed in the relevance perceived (i.e. among the categories).

**Expert groups differed in their perception of relevance**
We analyzed how different expert groups perceived relevance (raters: academic scientists vs. societal experts). Comparisons of the mean relevance of questions ($n_{BT} = 65; n_{TM} = 61; n_{AP} = 69$) revealed no significant differences for BT, with $t(64) = 1.217, p = .23$. In two of the three topics (TM and AP), the relevance perceived by two expert groups, academic scientists and societal experts, differed on average across all questions, with $t(60) = 2.766, p < .01$ for TM, and $t(68) = 7.114; p < .001$ for AP (see Figures 2a–c [1]).

**Questions with common authors were perceived as more relevant**
We tested whose questions were perceived as being relevant (authors: those of academic scientists vs. societal experts vs. both). Descriptive differences in the perceived relevance depended on the contributor of the question during the first round of Delphi (academic scientists or societal experts or both in cases of duplicate questions), with questions contributed by experts from both groups being more relevant (see Figures 2a–c [2]). However, there were no statistically significant differences in the mean relevance, with $F(2;62) = .833, p = .44$ for BT, $F(2;59) = 2.358, p = .10$ for TM, and $F(2;67) = 2.607, p = .08$ for AP.

**Perspective influenced perceived relevance of questions**
We analyzed how relevance differed between perspectives (relevant to the perspective: of a sender vs. a recipient vs. both). Questions on BT and TM assigned to both perspectives, that is, the dialogic model of communication between scientists and citizens, were perceived as more relevant only on a descriptive level, with $F(2;62) = 1.447, p = .24$ for BT, and $F(2;59) = 1.817, p = .17$ for TM (see Figures 2a–b [3]). Questions on AP were considered more relevant if they were shared by both perspectives, $F(2;67) = 3.773, p < .05$ for AP (see Figure 2c [3]).

![Figure 2a](image-url). Means and 95% Confidence Intervals on a 5-point Scale (5 = Relevant) for (1) Raters, (2) Authors, (3) Perspectives, (4) Consensus, and (5) Categories of Bats (BT). Note. scie. = academic scientists; pract. = societal experts; reci. = recipient; send. = sender; 2nd = second Delphi round; 3rd = third Delphi round; All other abbreviations refer to the categories’ names.
Consensus mattered

We tested whether relevance was higher when consensus was established in an earlier round (consensus: second vs. third round vs. none). For all three topics, an ANOVA revealed a similar trend across the consensus factor (consensus achieved in second round vs. third round vs. none), with $F(2;62) = 21.250, p < .01$ for BT, $F(2;59) = 4.710, p < .05$ for TM, and $F(2;67) = 30.190, p < .001$ for AP. Consensus was first established for more relevant questions in the second round, while consensus was reached on less relevant questions in the third round (see Figures 2a–c [4]). The mean relevance in the third round was comparable to the relevance of questions without a consensus.

Categories of questions revealed differences in perceived relevance

Last, we analyzed whether the overarching themes of the questions differed in the relevance perceived (i.e. among the categories of each topic). In two of the three topics (BT and TM), experts
perceived some categories as more relevant than others, with $F(9;54) = 6.125, p < .001$ for BT, and $F(7;60) = 2.821, p < .05$ for TM. For BT, post-hoc tests indicated that ecology, habitat, systematics, and the morphology of bats were perceived as being less relevant than the other categories of questions (all $ps < .05$). For TM, questions from the categories help desks, damages/damage prevention, and handling encounter situations were perceived as being more relevant than nutrition or urban habitat (all $ps < .05$). It was only in the case of AP that both expert groups did not perceive any differences between the categories of questions, with $F(10;58) = 1.327$ and $p = .238$.

Discussion

Our results confirm that both perspectives, those of citizens and of scientists, should be considered when deciding which questions a CS project aims to answer if the main goal of the CS project—that is to promote citizens’ learning about research topics—is to be met (e.g. Bonney et al., 2016). We were able to show that the experts assigned different levels of relevance to the questions depending on which perspective was considered, that of citizens or that of scientists. In principle, our results show that the academic scientists and societal experts differed significantly in their perception of relevance to the questions on TM as well as about AP and, on a descriptive level, also for questions on BT. Experts’ consensus on questions’ relevance, however, was more likely when the questions, for example, were related to both the sender and the recipient from a communicatory perspective. Our results indicate that questions on AP were considered more relevant if they were shared by both perspectives, and, on a descriptive level, also questions on BT and TM. We conclude that different knowledge about the topic area—indicated by the questions—is relevant to the perspectives of citizens and scientists. Expert consensus on relevance was reached if the two dimensions of relevance were considered, that is, structural and content-related aspects of the communication process between citizens and scientists influenced the perception of questions’ relevance (e.g. Bromme & Goldman, 2014; Bromme & Kienhues, 2017). The structural aspects are reflected in our results that are related to whose questions are perceived as relevant (i.e. those of academic scientists or societal experts) and to whom these questions are relevant (i.e. perspective of a sender or recipient). The content-related aspects are reflected in the different relevance depending on the thematic categories within the three topics. Thereby, our results contribute to previous research regarding the two dimensions of relevance (i.e. to whom and for what knowledge is relevant; Aikenhead, 2003; Mayoh & Knutton, 1997). First, the consensus achieved in our Delphi study indicates how structural aspects of the decision-making process in CS projects should move from a one-way deficit approach towards a dialogic account of citizens’ and scientists’ perspectives. The professional status of experts (academic scientists vs. societal experts) as well as the communicatory perspective (sender vs. recipient) indicated structural differences in the perception of relevance. Second, our analysis of the consensus indicated content-related differences in the relevance of questions; this highlights knowledge that could be enhanced by individual learning in CS projects from different fields (here: urban ecology on terrestrial mammals, bats, and air pollution) as it is useful from both citizens’ and scientists’ perspectives. Science-related but also non-science-related objectives in the questions’ themes also contributed to differences in the perception of relevant content.

One aim of this study was to achieve a consensus on the most relevant questions from scientists’ as well as citizens’ perspectives. These questions are indicators of topic-specific knowledge that is relevant to the individual learning of citizens in participatory CS projects. The second aim of this study was to conduct a detailed analysis of consensual questions to reveal structural as well as content-related differences. Our results reinforce previous findings that scientists have difficulties in perceiving what topics are relevant to citizens (Besley et al., 2015; Peterman & Young, 2015). Our results also show that differences in the expert groups’ ratings on the relevance of questions depended on their professional status (i.e. academic scientists and societal experts) and, therefore,
our results align with previous findings on scientists’ abilities to perceive citizens’ attitudes and knowledge about scientific research topics: Although scientists estimate citizens’ knowledge adequately (Bromme et al., 2001), they often fail to consider attitudes and needs (Besley et al., 2015; Carr et al., 2017), which is crucial for effective knowledge transfer. Our results support previous suggestions to not only elicit scientists’ perspectives on topics from their research but also citizens’ perspectives when it comes to the transfer of scientific research to the public (Bonney et al., 2016; Seakins & Dillon, 2013). Furthermore, our results extend previous findings by showing that common perceptions of relevance were more likely to occur when structural and content-related differences were accounted for.

**Questions need to be adjusted to include both communicatory perspectives**

If topic-specific knowledge is to be promoted in CS projects, structural differences in the communication process between citizens and scientists have to be accounted for. Therefore, participatory CS projects should overcome the deficit model by taking citizens’ perspectives into account in the decision-making process of framing questions (Bonney et al., 2016; Haywood & Besley, 2014). To promote knowledge that is relevant to both perspectives, we suggest that two structural differences have to be accounted for when defining what is to be learned from CS projects: To whom do the questions matter: sender or recipient? and Who contributed the questions?

As our results on all three topics indicate, the communicatory perspective matters. At least on a descriptive level, experts perceived questions as relevant if they mattered from both perspectives, that of the sender and the recipient (see Figure 2a–c [3]). The results show that questions that focused either on the sender’s or on the recipient’s perspective did not belong to the questions considered most relevant (i.e. following the unidirectional dissemination model; Trench, 2008). In line with previous research on dialogic models of science communication, considering both perspectives is more likely to stimulate communication about knowledge relevant to citizens and scientists as these models are more audience-centric (Besley et al., 2015; Carr et al., 2017). Our findings reinforce previous suggestions that scientists need to know about their audience’s characteristics (i.e. knowledge, interest, and motivation) in order to communicate effectively (Besley et al., 2015). Although marked differences were only found for questions on air pollution, we suggest that the relevance of topics from both perspectives should be considered when communicating about scientific research.

A similarly descriptive pattern was present in the results on the contributors of questions, that is, whether an academic scientist or a societal expert or both formulated a question; common perceptions of relevance were more likely if the particular question was suggested by both expert groups. This finding supports earlier suggestions that citizens’ perspectives need to be included in the decision-making process about the questions that are to be answered within CS projects (Bonney et al., 2016). Furthermore, our finding extends previous research on shifting from dialogic communication models to participatory models of CS: They show that both citizens and scientists can actually contribute questions that are then commonly perceived as relevant. The previously described shift in communication models applies to CS projects because they become more learner-centric (compared to audience-centric: Besley et al., 2015). Such learner-centric CS projects integrate both perspectives and promote the individual learning processes of citizens with regard to knowledge that is relevant to them.

**Content-related differences in the perceived relevance**

CS projects aim to promote citizens’ learning about topic-specific knowledge that is relevant to the citizens’ questions. To do so, considerations on the relevance of knowledge have to account for content-related differences that consider another dimension of relevance, that is, what the knowledge is relevant for (e.g. everyday life, academic research; Mayoh & Knutton, 1997). We compared themes
from urban ecology that come from different disciplines within the natural sciences (i.e. biology as well as physics and chemistry) in order to generalize our findings on content-related differences beyond the respective discipline. Our comparisons of themes from urban ecology suggest that themes are the most relevant from citizens’ and scientists’ perspectives when they are highly problem-oriented, that is, pose a risk to people. Problem-oriented questions relate to situations that require not only scientific knowledge but also non-scientific knowledge for their solution (i.e. normative decisions; Bromme & Goldman, 2014). For example, on the one hand, questions on the protection of biodiversity in cities (BT6 and TM6) or on protection from air pollution (AP2 and AP7) require scientific knowledge on effective protection measures and conditions of urban biodiversity (e.g. McDonald et al., 2020; Stillfried et al., 2017). On the other hand, questions on the protection of biodiversity in cities or on the reduction of air pollution require knowledge about legal regulations and are normative. Health-related issues are another example: While it is crucial to understand what citizens want to know, the framing of messages is crucial to avoid opposite effects, as discussed, for instance, in the case of bats and rabies (Lu et al., 2016). Citizens’ searches for information are often problem-oriented (Besley, 2014; Hendriks et al., 2016) and relate not only to scientific knowledge but also to normative aspects (de Groot et al., 2013). Taking the consensus as a starting point, experts considered problem-oriented questions to be relevant from citizens’ as well as scientists’ perspectives. Citizens’ questions often concern novel research and, therefore, when searching the internet for information, different positions have to be weighed up by applying normative framing (Bromme & Kienhues, 2017). The topic of diseases is often highly relevant to individuals when they choose topics to learn about (Baram-Tsabari et al., 2009). Our findings add to the research by showing that this topic is also relevant to scientists and, therefore, might serve as a starting point to introduce current research on urban ecology into CS projects (e.g. distribution patterns of wild mammals in the city).

Limitations
The strength of the Delphi technique in relying on experts is also its weakness. In this study, the local specificity of the topics restricted the selection of experts. We consulted experts who were working with citizens from the respective cities. Although this Delphi study provides an analytical approach to the relevance of questions on overarching themes in urban ecology, some questions might differ with regard to the local context in which the questions are asked (e.g. whom to contact in the case of urban wildlife conflicts).

Furthermore, we asked societal experts with expertise in citizens’ most frequent questions about the topics at hand, who, therefore, represented the ‘have-cause-to-know science’ dimension of relevance (Aikenhead, 2003, p. 20). Although these societal experts have a more accurate perception of which knowledge is relevant to citizens than, for example, academic scientists, future research should also account for the ‘personal curiosity science’ dimension (Aikenhead, 2003, p. 22) by collecting citizens’ questions first hand (e.g. Baram-Tsabari et al., 2006). At this point, the relevance of the questions to citizens could be further tested through the approach of communicative validation (Flick, 2009) by discussing both the questions themselves and their categorization with citizens.

Implications
When designing CS projects, scientists should consider integrating citizens’ perspectives into the questions examined by the project if the project goal that citizens learn about the research topic is to be met (Bonney et al., 2016). We suggest that designers of CS projects integrate the perspective of citizens when choosing the questions that the project aims to examine. Citizens’ questions are not necessarily limited to canonical knowledge about a subject (e.g. systematics and morphology for bats); they can also comprise normative questions (e.g. how safe it is to eat vegetables from the garden despite the presence of zoonosis). In prior research, questions that citizens asked scientists were primarily about problem-oriented topics when the questions were asked spontaneously. Questions were
more likely to be about academic topics when asked in a formal learning context (Baram-Tsabari et al., 2006). Individuals do not strive for expertise in a scientific discipline; they strive to answer questions that they perceive as being relevant to them and their everyday lives (Falk et al., 2007).

**Conclusion**

CS projects are increasingly changing into opportunities for participating citizens to learn, and they explicitly aim to promote citizens’ topic-specific knowledge as a learning outcome. We conducted a Delphi study to identify knowledge that is relevant to both citizens and scientists and to provide evidence for structural and content-related differences in the relevance perceived. We used questions as indicators of knowledge that might be relevant enough to learn about. Our study extends previous research that showed that CS projects do not promote knowledge relevant to citizens if they decide which knowledge is relevant only by considering general indicators of scientific literacy (Haywood & Besley, 2014; Stocklmayer & Bryant, 2012). We present evidence from three different topics speaking against the assumption that the knowledge that scientists consider relevant concurs with the knowledge that citizens consider relevant. Moreover, we show that the perceptions of which knowledge is considered relevant were more likely to agree when we accounted for structural and content-related differences, that is, to whom knowledge is relevant and for what knowledge is relevant. The experts were more likely to agree in their perception of the relevance of the knowledge when questions about it concerned both senders and recipients, or when the knowledge concerned practical problems that cannot be answered only by taking a scientific approach. Eliciting experts’ perceptions of relevance from two perspectives (i.e. academic scientists from the natural sciences and societal experts from state authorities and non-governmental institutions) provided us with an analytical lens regarding the relevance of topics for CS in urban ecology. If future CS projects are to achieve the goal of citizens’ learning about topic-specific knowledge, we suggest that informed decision-making processes that integrate both the perspective of citizens and that of scientists into the project design are necessary.

**Acknowledgments**

The authors would like to thank all of the experts who participated in the WTimpact project and all of the members of the WTimpact consortium.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by the German Federal Ministry of Education and Research (BMBF) [grant numbers 01O1725; 01O1727].

**Ethical statement**

The authors confirm all subjects have provided appropriate informed consent and details on how this was obtained are detailed in the manuscript.

**Declaration of conflicting interests**

The authors declare that there is no conflict of interest.
**Author note**

Till Bruckermann, former: IPN • Leibniz Institute for Science and Mathematics Education, Kiel, Germany; current: Institute of Education, Leibniz University Hannover, Hannover, Germany; Milena Stillfried, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany; Tanja M. Straka, former: Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany; current: Department of Ecology, Technische Universität Berlin, Berlin, Germany; Ute Harms, IPN • Leibniz Institute for Science and Mathematics Education, Kiel, Germany.

**ORCID**

Till Bruckermann https://orcid.org/0000-0002-8789-8276
Tanja M. Straka https://orcid.org/0000-0003-4118-4056
Ute Harms https://orcid.org/0000-0001-6284-9219

**References**

Aikenhead, G. S. (2003). *Review of research on humanistic perspectives in science curricula*. European Science Education Research Association (ESERA) conference, Nordwijkhout, Netherlands, 19–23 August 2003. Nordwijkhout, Netherlands.

Allum, N., Sturgis, P., Tabourazi, D., & Brunton-Smith, I. (2008). Science knowledge and attitudes across cultures: A meta-analysis. *Public Understanding of Science*, 17(1), 35–54. https://doi.org/10.1177/0963662506070159

Aristeidou, M., & Herodotou, C. (2020). Online citizen science: A systematic review of effects on learning and scientific literacy. *Citizen Science: Theory and Practice*, 5(1), 69. https://doi.org/10.5334/cstp.224

Baram-Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2006). Using questions sent to an ask-a-scientist site to identify children’s interests in science. *Science Education*, 90(6), 1050–1072. https://doi.org/10.1002/sce.20163

Baram-Tsabari, A., Sethi, R. J., Bry, L., & Yarden, A. (2009). Asking scientists: A decade of questions analyzed by age, gender, and country. *Science Education*, 93(1), 131–160. https://doi.org/10.1002/sce.20284

Bela, G., Peltola, T., Young, J. C., Balázs, B., Arpin, I., Pataki, G., Hauck, J., Kelemen, E., Kopperoinen, L., van Herzele, A., Keune, H., Hecker, S., Suškevičs, M., Roy, H. E., Itkonen, P., Külvik, M., László, M., Basnou, C., Pino, J., & Bonn, A. (2016). Learning and the transformative potential of citizen science. *Conservation Biology*, 30(5), 990–999. https://doi.org/10.1111/cobi.12762

Besley, J. C. (2014). Science and technology: Public attitudes and understanding. In National Science Board (Ed.), *Science and engineering indicators 2014* (7-1-7-53). National Science Foundation (NSB 14-01).

Besley, J. C., Dudo, A., & Storksdieck, M. (2015). Scientists’ views about communication training. *Journal of Research in Science Teaching*, 52(2), 199–220. https://doi.org/10.1002/tea.21186

Blanco-López, Á, España-Ramos, E., González-García, F. J., & Franco-Mariscal, A. J. (2015). Key aspects of scientific competence for citizenship: A delphi study of the expert community in Spain. *Journal of Research in Science Teaching*, 52(2), 164–198. https://doi.org/10.1002/tea.21188

Bonney, R. E., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T. B., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

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

Bromme, R., & Goldman, S. R. (2014). The public’s bounded understanding of science. *Educational Psychologist*, 49(2), 59–69. https://doi.org/10.1080/00461520.2014.921572

Bromme, R., & Kienhues, D. (2017). Gewissheit und Skepsis: Wissenschaftskommunikation als Forschungsthema der Psychologie. *Psychologische Rundschau*, 68(3), 167–171. https://doi.org/10.1026/0033-3042/a000359

Bromme, R., Rambow, R., & Nückles, M. (2001). Expertise and estimating what other people know: The influence of professional experience and type of knowledge. *Journal of Experimental Psychology: Applied*, 7(4), 317–330. https://doi.org/10.1037/1076-898X.7.4.317

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

Bruckermann, T., Greving, H., Schumann, A., Stillfried, M., Börner, K., Kimmig, S. E., Hagen, R., Brandt, M., & Harms, U. (2021). To know about science is to love it? Unraveling cause–effect relationships between knowledge and attitudes toward science in citizen science on urban wildlife ecology. *Journal of Research in Science Teaching*, 58(8), 1179–1202. https://doi.org/10.1002/tea.21697
Phillips, T. B., Porticella, N., Constas, M., & 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), 3. https://doi.org/10.5334/cstp.126

Ryder, J. (2001). Identifying science understanding for functional scientific literacy. *Studies in Science Education, 36*(1), 1–44. https://doi.org/10.1080/03057260108560166

Schüttler, S. G., Sorensen, A. E., Jordan, R. C., Cooper, C., & Shwartz, A. (2018). Bridging the nature gap: Can citizen science reverse the extinction of experience? *Frontiers in Ecology and the Environment, 16*(7), 405–411. https://doi.org/10.1002/fee.1826

Seakins, A. (2015). Revealing questions: What are learners asking about? In D. Corrigan, C. Buntting, J. Dillon, A. Jones, & R. Gunstone (Eds.), The future in learning science: What’s in it for the learner? (1st ed., pp. 245–262). Springer. https://doi.org/10.1007/978-3-319-16543-1_13

Seakins, A., & Dillon, J. (2013). Exploring research themes in public engagement within a natural history museum: A modified delphi approach. *International Journal of Science Education, Part B, 3*(1), 52–76. https://doi.org/10.1080/21548455.2012.753168

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

Stillfried, M., Gras, P., Börner, K., Göritz, F., Painer, J., Röllig, K., Wenzler, M., Hofer, H., Ortman, S., & Kramer-Schadt, S. (2017). Secrets of success in a landscape of fear: Urban wild boar adjust risk perception and tolerate disturbance. *Frontiers in Ecology and Evolution, 5*, 683. https://doi.org/10.3389/fevo.2017.00157

Stocklmayer, S. M., & Bryant, C. (2012). Science and the public—what should people know? *International Journal of Science Education, Part B, 2*(1), 81–101. https://doi.org/10.1080/09500693.2010.543186

Stocklmayer, S. M., Rennie, L. J., & Gilbert, J. K. (2010). The roles of the formal and informal sectors in the provision of effective science education. *Studies in Science Education, 46*(1), 1–44. https://doi.org/10.1080/03057260903562284

Stuckey, M., Hofstein, A., Mamlok-Naaman, R., & Eilks, I. (2013). The meaning of ‘relevance’ in science education and its implications for the science curriculum. *Studies in Science Education, 49*(1), 1–34. https://doi.org/10.1080/03057267.2013.802463

Trench, B. (2008). Towards an analytical framework of science communication models. In D. Cheng, M. Claessens, T. Gascoigne, J. Metcalfe, B. Schiele, & S. Shi (Eds.), *Communicating science in social contexts: New models, new practices* (pp. 119–135). Springer. https://doi.org/10.1007/978-1-4020-8598-7_7

Wals, A. E. J., Brody, M., Dillon, J., & Stevenson, R. B. (2014). Convergence between science and environmental education. *Science, 344*(6184), 583–584. https://doi.org/10.1126/science.1250515