Faculdade de Design, Tecnologia e Comunicação
Universidade Europeia

2023

CATALINA PÍA
LAGOS ROJAS

CREATIVE COMPUTING TOOLS FOR QUALITATIVE RESEARCH IN CITIZEN SCIENCE

Faculdade de Design, Tecnologia e Comunicação
Universidade Europeia

2023

CATALINA PÍA LAGOS ROJAS

CREATIVE COMPUTING TOOLS FOR QUALITATIVE RESEARCH IN CITIZEN SCIENCE

Dissertação apresentada ao IADE - Faculdade de Design, Tecnologia e Comunicação da Universidade Europeia, para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Computação Criativa e Inteligência Artificial realizada sob a orientação científica do Doutor Jacinto Paulo Simões Estima, Professor Auxiliar do Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra e do Doutor João Alfredo Fazendeiro Fernandes Dias, Professor Associado do IADE - Faculdade de Design, Tecnologia e Comunicação da Universidade Europeia.

Dedico este trabalho a Mónica e Rafael

agradecimentos

I would like to express my gratitude to my advisors, Professor Jacinto Estima and Professor João Dias, for their guidance and encouragement. Their expertise and dedication was integral to the successful completion of this thesis. Furthermore, I would like to acknowledge the faculty professors who have been part of this journey. Their insightful feedback, academic perspective, and support have significantly enriched this research process.
I wish to extend my sincere thanks to all the participants who generously gave their time and effort to partake in this study. Without their cooperation and contribution, this work would not have been possible.
Lastly, I am grateful to my family and friends, as their unwavering faith, understanding, and patience have been a constant source of strength during this journey.

palavras-chave computação criativa; ciência cidadã; investigação qualitativa; inquéritos

resumo A presente investigação explora o potencial da computação criativa para enriquecer a investigação qualitativa em ciência cidadã, centrando-se no envolvimento dos participantes. A principal questão de investigação é 'Pode a computação criativa ser utilizada para criar ou melhorar ferramentas para a investigação qualitativa em ciência cidadã, dando prioridade ao envolvimento dos participantes?' A hipótese sugerida é que métodos de computação criativa podem criar envolvimento por meio da interatividade, impactando a qualidade dos dados coletados em processos de pesquisa qualitativa. Foi realizado um conjunto de experiências para avaliar o efeito das generative data representations nas respostas abertas a inquéritos em linha e para compreender a interação dos participantes com elas, e a possível influência nas respostas. Os resultados confirmaram que esta abordagem pode, de facto, aumentar o envolvimento dos participantes, e melhorar o conteúdo das respostas. Este trabalho ilustra o potencial da aplicação da computação criativa aos métodos de investigação em ciência cidadã, proporcionando uma abordagem única à recolha de dados qualitativos. Estabelece uma plataforma para investigação futura e potenciais aplicações no âmbito de iniciativas de ciência cidadã.

keywords creative computing; citizen science; qualitative research; online surveys

abstract This research explores the potential of creative computing to enrich qualitative research in citizen science. The main research question is 'Can creative computing be used to create or improve tools for qualitative citizen science research, prioritizing participant engagement?' The hypothesis suggested is that creative computing methods can create engagement through interactivity, impacting the quality of data collected in qualitative research processes. A set of experiments was carried out to evaluate the effect of generative data representations on open-ended responses to online surveys and to understand participants' interaction with them, and the possible influence on responses.

The results confirmed that this approach can indeed increase participant engagement and improve the content of responses. This work illustrates the potential of applying creative computing to citizen science research methods, providing a unique approach to collecting qualitative data. It establishes a platform for future research and potential applications within citizen science initiatives.

Contents

Acronyms

BPM business process management. 9

HCI Human-Computer Interaction. 17, 18, 29, 33

IDE integrated development environment. 30

IxD Interaction Design. 20, 36

PAR Participatory Action Research. 15

QR quick response. 23, 29, 47

SDGs Sustainable Development Goals. 2

UN United Nations. 2

UX User Experience. 20

1 Introduction

As the world faces complex scientific challenges, from climate change to pandemics, the traditional boundaries of who participates in the creation of scientific knowledge are increasingly challenged. Citizen Science is revolutionizing how people and communities relate to science, serving as means for democratizing the scientific field, as it expands its reach to non-scientist citizens. However, several challenges arise from this change of paradigm, such as engaging participants throughout the whole research process and the communication of findings to the involved communities. Therefore, it is key to develop tools to support the expansion of citizen science, so that its scope can be extended to virtually all scientific areas.

This thesis undertakes a novel exploration at the intersection between creative computing and citizen science, a junction full of unexplored potential. This introductory chapter discusses the emerging importance of citizen science, highlighting its potential in democratizing science and facilitating community engagement. The focus then is narrowed down to the role that creative computing can play in this context, setting the stage for the research questions and objectives. Finally, a document structure overview is provided.

1.1 Citizen Science

Citizen science is an increasingly influential method for conducting and democratizing science, while promoting universal and equitable access to scientific data and information (de Sherbinin et al., 2021). It has become a powerful driver of contemporary scientific research. By integrating citizens into the processes of scientific investigation, citizen science brings science closer to the community, fostering active dialogue and engagement in decision-making processes (Constant & Roberts, 2017). This approach not only augments the body of scientific knowledge, but also empowers non-professional researchers and communities (Amirrudin et al., 2021), offering them opportunities for learning about policies, stimulating debates, and facilitating collaborative knowledge sharing (Williams, 2020).

The influence of citizen science research has been growing over the recent years, with thousands of individuals across various countries increasingly participating in these initiatives (Tauginienė et al., 2020). A testament to this rising trend is the European

Union Prize for Citizen Science, an initiative designed to support citizen science initiatives from a diverse range that contribute to Open Science (Ars Electronica, 2023). This prize, funded by the European Commission and awarded annually by Ars Electronica, signals the rising recognition of citizen science within the wider scientific community.

These growing efforts to cultivate citizen science projects are due to the multiple advantages offered by this approach. Grounded in the principles of participation, citizen science encourages volunteers to become engaged in solving scientific challenges (Haklay, 2013), fostering learning about the issue at hand and the scientific process at large. Furthermore, the data generated by citizen science initiatives are gaining recognition as invaluable resources for scientists, governmental bodies, and organizations, highlighting the immense potential of the approach (de Sherbinin et al., 2021).

1.1.1 The Value of Data

Citizen science holds substantial value in its capacity to generate data, the scale of which would be largely unattainable for individual researchers or research teams (Bonney et al., 2014). Enabled by the power of crowdsourcing, it provides scientists, governments, and organizations with a source of mass acquisition of data (Lee et al., 2020).

Citizen science is increasingly fueling scientific research and the pursuit of the 2030 Sustainable Development Agenda, particularly in biodiversity and pollution studies. Its crowdsourced data is not only employed in humanitarian activities by United Nations (UN) agencies but also in monitoring Sustainable Development Goals (SDGs) (de Sherbinin et al., 2021).

Moreover, the data generated in citizen science projects can facilitate active dialogue around issues relevant for a community, and engender engagement in decision-making processes (Constant & Roberts, 2017). Community-based research frameworks, in certain citizen science projects, serve to unite scientists and communities in addressing complex social issues (Lorenz, 2020). These projects can spotlight community concerns and serve as a tool for lobbying for better policies, potentially inspiring sustained and consistent data collection by government agencies on issues of importance to communities (de Sherbinin et al., 2021).

Although citizen science has been predominantly pursued within the natural sciences (Crain et al., 2014), citizen social science initiatives represent a significant opportunity to incorporate citizens into knowledge production, linking them with the policymaking processes and redefining the relationships between citizens, knowledge generation, and the state (Purdam, 2014). The facilitation of qualitative research processes can expand the impact of citizen science beyond the realm of natural sciences.

1.1.2 The Role of Qualitative Research

Qualitative research plays a fundamental role in improving our understanding of the context underlying numerical data. This approach seeks to generate data that is rich, nuanced, and comprehensive, allowing patterns and discoveries to surface through rigorous analysis (Barrett & Twycross, 2018). Unfortunately, it is frequently assumed that this approach competes with quantitative research, whereas in fact, both approaches complement each other (Silverman, 2021). Certain qualitative sensibilities combined with quantitative approaches result in more reliable, thorough, and ethical research processes and outcomes (Tanweer et al., 2021).

Qualitative research is particularly valuable when a complex, detailed understanding of an issue is required. This level of depth can only be achieved by studying individuals directly in their context, allowing their narratives to surface without being influenced by researchers' preconceptions (Creswell, 2017). By investigating why and how a phenomenon takes place, and understanding the meanings of concepts and features, qualitative research offers a rich exploration of the phenomena at hand (Longo, 2020).

1.1.3 Challenges in Ensuring Data Quality

In citizen science projects, ensuring data quality often presents a significant concern. The diversity of projects makes it challenging to assess whether the collected data meet the criteria for scientific research. Furthermore, the stakeholders involved can significantly influence the protocols' design and purpose (Balázs et al., 2021). Even within projects from the same scientific domain, diverse approaches and protocols exist, each establishing its own standards for data quality, and if the results are then generalized to the lowest common granularity, the resulting data might not meet the original quality thresholds (Balázs et al., 2021).

The issue of data quality becomes increasingly complex within the context of qualitative research. The use of qualitative methods—such as participant observations, in-depth interviews, and focus groups—demands a particular set of skills, including a well-developed ability to facilitate discussions and a nuanced understanding of data collection processes (Longo, 2020). The necessity of these skills points to an underlying com-plexity: the human element inherent in qualitative research can introduce variability and bias, making the task of maintaining data quality more challenging (Hanson & Marks, 1958).

The complexity of data quality in qualitative research extends to the process of data analysis. After data has been collected, it needs to be processed and understood. This typically involves methods such as coding or thematic analysis to identify and distill common themes (Vaughn & Turner, 2016). This step is not only time-consuming but also introduces another layer of potential subjectivity, as it requires researchers to make interpretive decisions about the data. Furthermore, the process of reducing qualitative data to identifiable themes or codes can, if not handled carefully, risk oversimplification, possibly diluting the richness and depth that qualitative data can offer (Guetterman et al., 2018). This poses yet another challenge to ensuring data quality in citizen science projects employing qualitative research methods.

1.1.4 The Power of Tools

In citizen science, tools such as apps and devices often serve as the principal interface between participants and the research process, enabling citizens to contribute to the scientific endeavor. The design of these tools and platforms has a considerable impact on various aspects of a project. For instance, it can influence participant motivation (Land-Zandstra et al., 2021), the frequency of contributions (Sauermann & Franzoni, 2015), and even the overall scope and impact of the project (Williams, 2020).

Mobile applications play a crucial role in many of these projects, serving as an extended sensor for users in quantitative research (Lemmens et al., 2021). These apps typically prioritize structured data input and precision due to the numerical, rigid nature of quantitative research. However, the demands of qualitative research are considerably different. As qualitative research seeks to explore meanings, interpretations, symbols, and relational processes, apps suited for qualitative research should support flexible, open-ended inputs and rich interactions among participants and researchers.

The citizen science research design process, which encompasses everything from formulating the research questions to producing evidence-based collective results, should be inclusive, flexible, and adaptive (Senabre Hidalgo et al., 2021). When developing tools to support these processes, usability, user interaction, and interoperability must be considered (Lemmens et al., 2021). However, it is essential to note that merely adapting existing quantitative solutions may not fully address the unique requirements of qualitative research.

The mediating role of technology in qualitative research adds an additional layer of complexity. As qualitative research relies on human interactions and interpretations, the design of digital tools must be sensitive to how technology can impact data collection processes. In addition, ethical and privacy considerations become even more important when personal stories and experiences form the core of the data. Creating effective tools for qualitative research in citizen science calls for thoughtful interaction design, tailored to the unique requirements and challenges of this research approach.

1.2 Research Background

The use of citizen science for public participation in knowledge creation has largely found its stronghold within the natural sciences. These domains often utilize quantitative methodologies that produce structured and numerically precise data, aligning well with conventional scientific research approaches. However, the necessity to expand the reach of citizen science, particularly into the realms of social sciences, prompts the incorporation of qualitative research methodologies and instruments.

Qualitative approaches, marked by their flexible, open-ended queries and deep, interactive engagements, present unique challenges when applied within a citizen science context. These challenges mainly manifest in data collection, participant engagement, and data analysis, where the demands differ considerably from those of conventional scientific research or even quantitative citizen science projects. Additionally, the digital tools typically employed in citizen science projects, designed predominantly with a quantitative approach, do not adequately cater to the needs of qualitative research.

1.2.1 Problem Statement

The challenges of the incorporation of qualitative research in citizen science are core of this study, as its main objective is to explore the potential of creative computing to generate or enhance tools for qualitative research in citizen science, thereby boosting participant engagement and enriching data collection and analysis.

Thus, the primary problem this study aims to address is how can the field of citizen science, traditionally skewed towards quantitative methodologies, effectively integrate and harness qualitative research approaches, especially via the application of creative computing? To address this issue, an exploration into the transformative potential of creative computing in citizen science is required, including the formulation and experimentation of innovative approaches to qualitative data collection and analysis.

1.2.2 Motivation

The driving force behind this study arises from the challenge and opportunity to harness the interdisciplinary potential of creative computing, linking fields such as software development, interaction design, and social sciences within the context of citizen science. This is rooted in the personal conviction that a creative computing approach can foster a convergence of disciplines that can fuel innovation and foster empowerment.

Fundamentally, this research is motivated by the principle that technologies should be created to serve the needs of our envisioned solutions, rather than constraining our solutions within the bounds of existing technologies. This resonates with the old proverb: 'If you want to go fast, go alone. If you want to go far, go together'; to achieve meaningful change and empowerment—an often overlooked byproduct of effective design decisions—it is essential to adopt an integrated, interdisciplinary perspective.

The design and development of the right tools can empower communities and organizations to take action toward more equitable and sustainable futures. This potential is exemplified in many citizen science projects that involve participants throughout the entire research process, returning the knowledge generated to the communities where it is most relevant. Hence, the motivation for this study lies in exploring how creative computing can contribute to such empowerment and change within citizen science, particularly through the integration of qualitative research methodologies.

1.2.3 Research Question and Objectives

The primary research question guiding this study is: **'Can creative computing be utilized to generate or augment tools for qualitative research in citizen science, prioritizing participant engagement?'**. Based on this question, the main hypothesis is that creative computing methods can enhance engagement throughout the qualitative research process, thus allowing for the creation of tools focus on participant's motivation, resulting in improved data collection processes.

To test this hypothesis and to answer the primary research question, the following objectives were set:

1. Design and develop creative computing tools specifically designed for qualitative data collection within citizen science projects.
2. Conduct experiments to evaluate the developed tools to assess the influence on participants' responses within an online survey.
3. Analyze and determine the effectiveness of the tools in improving the quality and quantity of collecting qualitative data, through key variables.
4. Evaluate the overall effectiveness of the developed tools in enhancing qualitative research in citizen science, from a user experience perspective.

1.3 Document Structure

The rest of this document is structured as follows:

Chapter 2: Theoretical Framework delves into the intricate elements of data-driven innovation, citizen science, online surveys, online interactivity, free text as data, and the representation of free text. The aim of this chapter is to establish a robust theoretical foundation for the subsequent empirical study.

Chapter 3: Related Work critically examines the current landscape of citizen science research, online surveys, creative computing, and alternative methods of data collection. This chapter illuminates advances in these areas, setting the stage for the proposed methodology.

Chapter 4: Methodology details the research approach, establishing why creative computing can be leveraged for this study. Generative data representations are introduced and analyzed in detail, to understand how are they generated, and their features. Finally, the research design for this study is presented.

Chapter 5: Experiments details the use case for this research, and the developed interfaces for it. This chapter explains in detail each experiment design, the data analysis approach, ethical considerations, and finally, how the experiments were implemented.

Chapter 6: Results presents the outcomes of the experiments conducted. Both quantitative and qualitative results are discussed, with data visualizations used to facilitate a comprehensive understanding. Furthermore, an in-depth interpretation of the results obtained is presented. This chapter includes the evaluation of the hypothesis, and contextualization of the results within the broader research framework.

Chapter 7: Conclusions wraps up the study by highlighting its principal findings and their implications. The research question and objective are addressed, and a discussion of potential directions for future research is included, building on the insights obtained from this study.

2 Theoretical Framework

Having outlined the overall context, detailed components of the problem, emerging challenges, and project objectives, this chapter aims to provide a comprehensive review of the relevant literature. The goal here is to establish a foundation for understanding the interplay between the various areas under study, such as data-driven innovation, citizen science, online surveys, and online interactivity.

2.1 Data-driven innovation

We currently live in an age characterized by an unprecedented abundance of data. The term *Big Data* is widely used across various sectors, from meteorology, complex physics simulations, environmental research, finance and business, to healthcare (Yin & Kaynak, 2015). It denotes an extensive volume of data featuring diverse and intricate structures that conventional data management methods and procedures cannot handle efficiently (Rabhi et al., 2019). The concept is often considered broad, and thus, it lacks conceptual clarity (Kitchin & McArdle, 2016). In essence, Big Data refers to the handling and analysis of massive datasets and is now becoming integral to modern science and business (Diebold, 2012).

The digital information constituting Big Data originates from multiple sources. Data can be automated—continuously recorded at high speeds across multiple sources—or human-mediated, with individuals creating records (i.e., administrative records in public services). While human-mediated data also ensures a steady stream, its volume and velocity are considerably lower than automated systems (Kitchin & McArdle, 2016).

Data-driven innovation depends on the quantity and quality of available data. Although it can be conducted without Big Data, the latter undoubtedly brings additional dimensions to the process by expanding the available resources. Access to relevant and accurate data can streamline decision-making processes. In business, Big Data introduces new data sources to business process management (BPM) and presents opportunities for more profound analysis and potential improvements (Rabhi et al., 2019).

The use of Big Data requires a shift in analysts' workflows from isolated 'pools' of data to 'fast-flowing streams', it mandates a more continuous strategy for data collection, analysis, and action (H. Davenport, 2014). This permeates virtually all business sectors, from human resource management, encompassing recruitment processes, training, and career management (Rabhi et al., 2019), to sports organizations aiming to enhance player performance (Yin & Kaynak, 2015).

However, data abundance does not equate to an abundance of knowledge. One of the challenges associated with using Big Data to generate knowledge is the risk of misinterpretation. Since the datasets are often voluminous, they fulfill the sample size requirements of many statistical procedures, which can confer a false sense of certainty. However, these datasets should not be employed uncritically, as they often carry significant biases (McFarland & McFarland, 2015).

Traditional data challenges, such as noise, outliers, incomplete, and inconsistent data, are magnified in the realm of Big Data (Tsai et al., 2015). These issues can pose considerable risks, particularly when coupled with data from online sources, which are typically not derived from statistically rigorous experiments. Additionally, these sources often contain various forms of biases, potentially leading analysts towards 'precisely inaccurate' results (McFarland & McFarland, 2015).

Big Data is not confined to quantitative data; it also comprises qualitative data such as social media posts, images, call center recordings, and product reviews, among other forms. Tasks like sentiment analysis can be performed to extract qualitative insights from these data sources, for instance, to determine public sentiment towards a brand over time.

2.1.1 Complementing with qualitative research

Qualitative research is particularly beneficial when nuanced insights are required. It's important to note that qualitative research encompasses more than just the use of qualitative data; it represents an entire research approach. In his book *Qualitative Inquiry and Research Design*, Creswell (2017) argues that qualitative research is used when theories are partial or inadequate. Occasionally, quantitative measures and statistical analyses fail to fully address or represent the complexity of the research subject. For instance, quantitative methods may prove inadequate when exploring nuanced areas such as gender inequalities, racial disparities, and individual differences that may not be well represented in the data.

Unfortunately, qualitative research is often associated with misleading assumptions, such as the belief that it competes with quantitative research. In reality, it complements quantitative research by delving into the 'black box' of how social phenomena form in real-time (Silverman, 2021). The focus of qualitative research is to investigate why and how a phenomenon takes place, by understanding the meanings of concepts, definitions, and features, rather than measuring their occurrences or frequencies (Longo, 2020).

As stated, qualitative research is crucial when a complex and detailed understanding of an issue is needed. This depth can only be achieved by studying individuals directly in their contexts, allowing their stories to emerge independently from any preconceived theories researchers might hold (Creswell, 2017). To yield these insights, qualitative research requires data that is comprehensive, rich, and nuanced, facilitating the emergence of patterns and discoveries through rigorous analysis (Barrett & Twycross, 2018). The distinctions between qualitative and quantitative research approaches in terms of data can be summarized as follows:

Data collection: Quantitative research typically involves the collection of numerical data, whereas qualitative research involves the collection of non-numerical data, such as observations and insights from interviews and focus groups.

Sampling: Quantitative research usually requires a larger sample size and tends to use random sampling techniques, while qualitative research may use a smaller sample size and is more likely to employ non-probability sampling techniques.

Data analysis: Quantitative data is generally analyzed using statistical methods, while qualitative data is analyzed using methods such as content analysis, discourse analysis, and thematic analysis.

Research questions: Quantitative research tends to pose questions that can be answered with numerical data, while qualitative research is more likely to ask questions that require non-numerical data for responses.

Results: Quantitative research typically results in numerical data and statistical analyses, while qualitative research results in detailed descriptions and interpretations of the data.

2.1.2 Changes in qualitative data collection methods

In discussions about qualitative research methods, semi-structured methods are often referred. Those include participant observations, in-depth interviews and focus groups (Longo, 2020). However, these methods have evolved over time due to changes in the digital landscape, such as the advent of the Internet and social media platforms. These digital tools have enabled researchers to expand traditional methods (Twis et al., 2020). For instance, video conferencing software has made remote interviews possible, eliminating the need for the interviewer and participants to share the same physical location. While online interviews are essentially an extension of traditional interview methods, they are increasingly viewed as a new method (Thunberg & Arnell, 2022).

The adoption of digital tools extends beyond research interviews to include focus group methods (Davies et al., 2020). Remote participation not only offers cost savings and improved accessibility, but it also facilitates discussions on sensitive topics. However, researchers have noted that online responses are typically shorter and more concise, with fewer opportunities for follow-up questions compared to in-person sessions (Davies et al., 2020).

Online surveys are often criticized for their methodological limitations, because the target population is variable and often not possible to characterize, thus possibly biased, which can make findings not generalizable and misleading (Andrade, 2020). Yet, they are being used with increased frequency (Couper, 2011), and can be useful for qualitative research, allowing to increase the outreach of qualitative in-person surveys.

In their simplest form, online qualitative surveys can consist of a series of open-ended questions on a specific topic. As respondents are encouraged to use their own words rather than choosing from pre-determined answers, these surveys can yield richer and more complex information (Clarke & Braun, 2013). However, there is no one-size-fits-all solution for data collection. Different research contexts and objectives may need to use different methods and implementation procedures (Stern et al., 2014).

2.1.3 A paradigm shift in data collection

The advent of the internet has broadened the scope of data collection methods, but the most recent years have witnessed an even more significant paradigm shift with the adoption of crowdsourcing techniques. Crowdsourcing is a method of generating information in which individuals or organizations obtain goods and services from a large, relatively open, and often rapidly evolving group (Charvát & Kepka, 2021). As Brabham (2008) defines it, crowdsourcing is an online distributed problem-solving and production model that meets specific organizational goals and shares power or agency between the crowd and the organization.

This concept can also be considered a community research method, in which a large group of people contribute data on a topic, either actively or through passive data collection (English et al., 2018). It is often purposed, collaborative, and co-designed to achieve a specific outcome (Lee et al., 2020).

When crowdsourcing is employed as a research method, non-scientist citizens who voluntarily participate are included in scientific research. This approach significantly deviates from traditional in-lab research paradigms, as the scale achievable with citizen participation is beyond the reach of individual researchers or research teams (Bonney et al., 2014). In essence, crowdsourcing enables the mass acquisition of data (Lee et al., 2020), opening up new opportunities for research.

2.2 Citizen science

The advent of new research methodologies offers multiple benefits to citizens, allowing them not just to learn about science but also to participate in it actively. This innovative approach also benefits scientists by providing them with a different data collection and analysis framework. Citizen science involves the active participation of non-scientists who assist in collecting or analyzing data as part of a researcher-led project (Gura, 2013), effectively bridging the gap between society and scientific research. Citizen science is a flexible concept (Gold, 2022), with its definitions varying across regions, scientific fields, and types of engagement due to its widespread application in numerous research initiatives (Campos et al., 2021).

Participants in citizen science can engage in various aspects of the research process, from project design and hypothesis formation, to data collection, classification, and analysis, to the communication of findings (West & Pateman, 2016). Therefore, volunteerism is central to the understanding of citizen science (Haklay et al., 2021).

This field is receiving increasing attention (Sauermann & Franzoni, 2015). In Europe, the European Citizen Science Association (ECSA) was established in 2013 to encourage the growth of citizen science in Europe, and to support the general public's participation in research processes – spanning science, social science, humanities, and the arts (European Citizen Science Association (ECSA), n.d.). It operates as a community of practice that supports these initiatives (Albert et al., 2021). Furthermore, last year, the Citizen Science Global Partnership (GSP) was initiated to foster international collaboration for projects and international organizations (Global Citizen Science Partnership, 2022).

2.2.1 Benefits

Citizen science offers a wealth of scientific and societal opportunities across Europe and beyond. On the one hand, individuals can engage in learning and personal development through their contributions to scientific knowledge. On the other hand, communities can strengthen their collaboration networks (Albert et al., 2021), and actively engage in dialogue and decision-making processes (Constant & Roberts, 2017). However, to ensure equitable participation, it is imperative to address the uneven access to funding and infrastructures (Albert et al., 2021).

Citizen involvement in research projects varies widely. Participants often perform tasks such as data collection, image coding, or document transcription (Sauermann & Franzoni, 2015), but if they wish, they should be allowed to participate in multiple stages, from developing the research question to communicating the results (Gold, 2022).

To classify participation levels, Bonney et al. (2014) proposed a widely used framework consisting of three project categories: contributory, collaborative, and co-created. In contributory projects, scientists design the project, and participants contribute to data collection and analysis. Collaborative projects enable participants to adjust protocols, draw conclusions, and even propose new research directions. Co-created projects are the most inclusive, allowing participants to engage in all stages of the scientific process, as these projects are collaboratively designed and developed.

While citizen science has primarily been implemented within the natural sciences (Crain et al., 2014), its reach extends into the social sciences, contributing not only to the enhancement of scientific knowledge but also to the empowerment of non-professional researchers and communities (Amirrudin et al., 2021). Citizen social science offers a platform for citizens to participate in knowledge production, thus fostering their engagement in the policymaking process and reshaping the interactions between citizens, knowledge processes, and the state (Purdam, 2014). Essentially, citizen social science investigates social issues as part of citizen science, using a community-based research framework to unite social scientists and communities in addressing complex social issues (Lorenz, 2020).

As with citizen science, citizen social science constructs datasets through the participation of wider, interested, and self-selected communities. This approach often aims to counterbalance data collected, controlled, presented, and sometimes overlooked by governments and institutions, as exemplified by Participatory Action Research (PAR) in the mid-1980s (Williams, 2020). Notably, citizen science and PAR share important similarities, as both are umbrella terms encompassing various related but subtly distinct concepts such as crowd science, participatory science, action science, and community-based research (Lorenz, 2020).

In particular, citizen social science embraces the diversity of scientific disciplines and human societies. Consequently, it strives to be both scientifically rigorous and community beneficial by integrating diverse academic and practical knowledge systems (Lorenz, 2020). Balancing these objectives can be challenging, given that community-based research often advocates for or against changes, whereas academic research primarily aims to produce generalizable knowledge (Amirrudin et al., 2021).

2.2.2 Challenges

Throughout the literature reviewed, three main challenges arise: (i) participant motivation and engagement, (ii) effective participant learning, and (iii) data quality.

Motivation. Participant motivation is critical to the success of citizen science projects (Land-Zandstra et al., 2021). Understanding participant motivation can help engage citizens effectively by aligning projects with their interests and experiences. Studies suggest participants are motivated by a mix of intrinsic factors, such as contributing to scientific knowledge or personal interest in the topic, and extrinsic factors, such as social interaction within a community of like-minded people (Hart et al., 2022; Land-Zandstra et al., 2021; Williams, 2020). However, sustaining participation over time remains a significant challenge, as engagement tends to decline after an initial period of activity (Campos et al., 2021; Sauermann & Franzoni, 2015).

Learning. Citizen science projects are often educational and can improve data literacy (Williams, 2020). Participants typically expect to learn more about the project topic, making learning opportunities and timely communication of project discoveries essential (Celino et al., 2021; Land-Zandstra et al., 2021). Unfortunately, the visibility of citizen science efforts is often limited, threatening their long-term viability (Lee et al., 2020). To enhance project success, there should be a focus on showcasing project achievements, impacts, and data accessibility (Druschke & Seltzer, 2012).

Data quality. The quality of data collected in citizen science projects can be a controversial issue. The criteria for scientific research may vary based on the project and the stakeholders involved, who may have different perspectives on data reliability, bias, and relevance (Balázs et al., 2021). Some projects prioritize the act of data collection as a learning tool over accuracy, using any inaccuracies as an opportunity to teach participants critical data analysis (Williams, 2020). However, to ensure data quality, a well-considered data collection and analysis plan should be established at the beginning of the project (Balázs et al., 2021).

2.3 Tools in Citizen Science

Citizen science projects use a variety of tools to facilitate data collection and enable active participation. These tools range from conventional research approaches, such as interviews, surveys, and focus groups, to novel digital technologies. In citizen science projects, all of these tools should be tailored to the unique, decentralized and participatory nature of it.

2.3.1 Qualitative Data Collection Tools

Qualitative data collection tools provide researchers with in-depth insights into the experiences and perspectives of citizen scientists. These tools, which traditionally include semi-structured methods, such as participant observations, in-depth interviews, and focus groups, are often used in qualitative research (Longo, 2020), allowing researchers to gather non-numerical data that can be analysed to reveal patterns, themes, and meanings.

In the realm of citizen science, interviews afford a significant degree of flexibility, allowing researchers to adapt to the needs and preferences of individual participants. The semi-structured, open-ended approach that is often favored in these interviews allows researchers to delve deeply into participants' experiences, perceptions, and motivations.

This approach, however, requires a high degree of skill and sensitivity on the part of the interviewer, and it can be time-consuming.

Surveys and questionnaires, on the other hand, offer a more streamlined and efficient approach to data collection. These tools, which can be distributed and completed digitally, allow researchers to gather data from a large number of participants. The inclusion of open-ended questions in these surveys and questionnaires provides an opportunity for qualitative data collection. However, researchers must be mindful of potential biases that can result from self-selection and self-reporting among participants (Andrade, 2020).

However, online data collection strategies are constantly expanding, resulting in many variations of traditional research methods. Researchers are now resorting to online questionnaires, online forums, Facebook, websites, blogs, e-mail, online focus group, Twitter, chats, and YouTube (De Oliveira Salvador et al., 2020).

2.3.2 Digital technologies

Digital technologies are revolutionizing citizen science by opening up new avenues for data collection, participant engagement, and project dissemination. These technologies range from mobile apps, facilitating on-the-spot data collection, to online platforms connecting citizen scientists globally.

These technologies are not merely enhancing data transmission and accessibility, but also fostering novel educational paradigms within citizen science. By engaging citizens directly in comprehensive scientific processes such as data collection, analysis, interpretation, and discussion, digital technologies democratize scientific discovery, thus fostering an inclusive scientific culture (Bruckermann et al., 2022).

In this realm, Human-Computer Interaction (HCI), a field dedicated to optimizing the interaction between users and computer technology, emerges as highly relevant. HCI is an essential element of computer science, it involves several disciplines (Grudin, 2005), and its knowledge can guide the creation of more effective, efficient, and engaging tools.

However, making appropriate technological decisions for a project can be a complex process. Selecting or developing the right tools requires careful consideration of project goals, intended users, type of data to be collected, and available resources. Moreover, potential high costs of technological investments necessitate a cost-benefit analysis.

Though HCI has historically focused on micro-level issues like usability and user satisfaction, its scope has expanded to encompass macro-level concerns, such as social, cultural, and ethical implications of technology use (Preece, 2016). This expansion becomes particularly significant within the diverse and inclusive context of citizen science.

There exists no universal framework for integrating digital technologies into citizen science projects; each project's unique needs may demand different toolsets and strategies. However, fundamental principles of HCI design like user-centered design, accessibility, and usability can inform tool selection and implementation.

Lastly, it is noteworthy that digital technologies can enhance traditional qualitative research tools' capabilities. For instance, online surveys can reach more diverse samples, and digital audio or video recordings can offer ease of storage, sharing, and analysis.

2.4 Online Surveys for Qualitative Research

While semi-structured methods such as participant observations, in-depth interviews, and focus groups are often used in qualitative research (Longo, 2020), their application in citizen science can be challenging due to the requirement of experienced researchers for unbiased results (Creswell, 2017). An alternative and suitable method for such contexts is the use of online surveys. Non-mediated and asynchronous, online surveys offer participants an avenue for qualitative data collection without necessitating extensive training or overcoming participation barriers.

2.4.1 Benefits and Limitations

Although online surveys are globally recognized for their role in quantitative research (Wu et al., 2022), their efficacy in generating robust qualitative insights is often debated among researchers. Critics argue that the data derived from online surveys may lack sufficient depth for producing strong, stand-alone insights (LaDonna et al., 2017), and may potentially introduce biases that cannot be fully addressed or assessed (Andrade, 2020).

Nonetheless, online surveys are valuable for qualitative research due to their flexibility, breadth of applicable research questions (Braun et al., 2021), and potential to engage a large, geographically dispersed sample (Braun et al., 2017; Menon & Muraleedharan, 2020). The inclusion of open-ended questions, or free-text responses, allows participants to express themselves in a more authentic way, leading to the generation of rich and intricate data that can better represent their subjective experiences, narratives, and viewpoints (Clarke & Braun, 2013). Even if individual responses are brief, the data accumulated through these surveys can deepen the understanding of specific issues (Braun et al., 2021).

In the context of citizen science projects, qualitative online surveys can offer valuable insights during exploratory phases, with free-text response analysis providing a preliminary understanding to direct further research (LaDonna et al., 2017). However, challenges such as low response rates and poor response quality are notable. Compared to traditional research methods (e.g., postal, telephone, face-to-face), web surveys often yield lower response rates due to survey fatigue, competing demands, and privacy concerns (Daikeler et al., 2020; Menon & Muraleedharan, 2020).

2.4.2 Survey Design

Response quality is influenced by the device used to complete the survey, with responses tending to be shorter on mobile surveys compared to desktop (Mavletova, 2013). This discrepancy could be due to survey designs often replicating paper questionnaires, leading to suboptimal user experiences on mobile devices (Toepoel et al., 2020). Additionally, mobile surveys typically take longer to complete (Couper & Peterson, 2017), as respondents may multitask more often, potentially affecting data quality (Antoun et al., 2017). These factors become significant in light of the increasing trend of online survey completion on smartphones (Tourangeau et al., 2018).

Moreover, survey length is inversely related to response length and response rates. Lengthier surveys may result in shorter responses and lower participation rates (Rolstad et al., 2011), emphasizing the importance of careful and concise survey design.

2.5 Enhancing Engagement Through Online Interactivity

Online interactivity encapsulates a variety of user engagement forms within the digital realm, from simple actions like clicking or filling forms to intricate activities like engaging in online games or content creation. This concept is central to fostering effective online communication and enhancing user engagement, with far-reaching implications in diverse fields like education, marketing, and research.

2.5.1 Gamification

Gamification, the application of game-design elements in non-gaming contexts, is frequently used to promote desired motivational, behavior and learning outcomes (Zainuddin, 2018). It can enhance user engagement, stimulate participation, and enrich learning experiences (Chou, 2015). This approach can be manifested through many mechanics and dynamics, such as point systems, badges, leaderboards, or quests, that can integrated within educational platforms, business applications, or citizen science projects (Krath et al., 2021).

Through its ability to tap into both intrinsic (i.e., mastery, curiosity, enjoyment) and extrinsic (i.e., points, badges, recognition) rewards, gamification can significantly stimulate participation (Zichermann & Cunningham, 2011). It can further promote social interactions and cooperative behavior within online communities.

However, the implementation of gamification requires thoughtful planning and design to ensure alignment with the intended users' motivations and the underlying objectives of the project or platform. Misguided gamification can have unintended side-effects, such as inadvertently promoting competition over collaboration, or placing undue focus on rewards over the inherent value of the activity (Almeida et al., 2021).

2.5.2 Interaction Design

Interaction Design (IxD) is a specialized field concentrating on the creation of meaningful and engaging interactions between users and digital interfaces (Interaction Design Foundation, n.d.). IxD aims to understand user needs and behaviors, to create useful, desirable, and accessible designs (Moggridge, 2003).

In the domain of online interactivity,IxD covers various elements, including user interface design, usability, accessibility, and overall User Experience (UX). The expertise is of prime significance in web design, application design, and any digital environment where user interaction is fundamental.

Well-crafted interaction design can make digital platforms more user-friendly and captivating, leading to greater user satisfaction and increased engagement levels (Goodman et al., 2011). This requires a cyclical process of user research, prototyping, testing, and iterative refinement (Preece, 2016).

2.5.3 Creative Computing

Creative computing can be understood as the innovative use of computer technology and coding to facilitate creative outcomes like art, music, interactive stories, or games. In other words, the action of computing but promoting human creativity (Yang & Zhang, 2016). It provides a medium for creative expression, an innovative tool for learning and teaching (especially in STEM education), or a methodology for engaging audiences through interactive experiences.

In the frame of online interactivity, creative computing can enhance user engagement through interactive visual and auditory elements, or interactive storytelling. Interactive data visualizations, for instance, can simplify complex data representation, making it more understandable and engaging. Likewise, interactive storytelling can offer immersive experiences that stimulate active participation (Zhang et al., 2019).

Despite these benefits, the application of creative computing requires mindful consideration of usability and accessibility to ensure that the creative aspects augment, rather than impair, the user experience. Moreover, acknowledging the diversity of users' skills, experiences, and cultural backgrounds is imperative to designing inclusive and accessible creative experiences

2.6 Summary

The theoretical framework presented in this chapter discussed the potential and challenges of citizen science, along with its applications and tools. Citizen science offers a pathway to democratize science by involving non-specialists in data collection and analysis, providing broader data collection and stimulating public scientific discourse.

Key challenges in citizen science include maintaining participant motivation, ensuring effective learning, and securing data quality. These are addressed by understanding participants' motivations, providing learning opportunities, and devising strategies to facilitate data collection.

Tools in citizen science range from traditional data collection methods to innovative digital technologies. These tools, customized to the participatory nature of citizen science, can enhance data collection, boost engagement, and promote an inclusive scientific culture.

Online surveys offer a pragmatic and extensive method for data collection in citizen science. Despite limitations such as low response rates and mobile device response constraints, well-designed online surveys can elicit valuable insights.

Strategies for online interactivity, like gamification, interaction design, and creative computing, play a significant role in enhancing user engagement. By aligning these strategies with users' motivations and project objectives, participation can be stimulated, user experience improved, and inclusive platforms created.

3 Related Work

This chapter presents an examination of diverse advances in the fields of data collection and citizen science. The exploration ranges from traditional techniques such as online surveys, to the utilization of exhibitions as tools for data collection, and the innovative domain of creative computing. The focus lies not only in understanding these advances but also in discerning the challenges intrinsic to their implementation and the strategies devised to navigate them. Drawing from a series of successful projects, the aim is to illustrate their potential for citizen science research. The role and potential of creative computing is scrutinized as a way to collect, engage, visualize, and disseminate information.

3.1 Online Surveys

As previously discussed, online surveys are a valuable tool in data collection across various research fields. They offer a cost-effective and efficient way of collecting information from a large and geographically diverse sample. They are non-mediated and asynchronous, which is beneficial for citizen science. However, they present a set of challenges, particularly regarding response rates, survey design, and participant engagement.

3.1.1 Response Rates

One of the long-standing issues with online surveys is the often low response rate. Numerous strategies to improve this metric have been studied, including preliminary contact with participants, shorter survey lengths, incentives, and reminder systems (Edwards et al., 2009). Other tactics include personalizing survey materials and using quick response (QR) codes on various communication channels for faster survey access (Harrison et al., 2019). However, not all strategies have been successful in their attempts. For instance, a study that incorporated tailored videos for participants, aiming to provide non-material incentives as a way to increase response rates, but it led to a decrease in them (Kalleitner et al., 2020).

3.1.2 Survey Design

In terms of survey design, there are many factors that can affect the quantity and quality of the gathered answers, such as the specific device it was designed for (Couper, 2011), survey length, presentation and content (Fan & Yan, 2010), and even the way questions are phrased. Along these lines, researchers have been working from different approaches on these issues. For instance, a study assessed the impact of a web survey design modeled as a messaging app, so it would mimic natural conversation (Toepoel et al., 2020). The study found that while the messaging-style survey resulted in shorter responses and slightly longer completion time, the quality of the data collected was comparable to the traditional survey.

When comparing face-to-face surveys with online self-administered ones, the former often generate higher response rates and richer data (Christensen et al., 2014; Daikeler et al., 2020). This disparity can be attributed to the fact that online surveys are increasingly answered on mobile devices, a potentially multitasking environment. This can translate into higher cognitive loads involved in answering them and the likelihood of respondents taking shortcuts to minimize these loads (Couper & Peterson, 2017; Heerwegh & Loosveldt, 2008).

3.1.3 Gamification Strategies

Efforts to engage participants through perceived fun and incentives have led to the exploration of gamification techniques in online surveys. In one study, researchers developed a mini-game-based survey, where each type of question was attributed to a different minigame. One of the mini-games proposed can be seen in Fig.1.

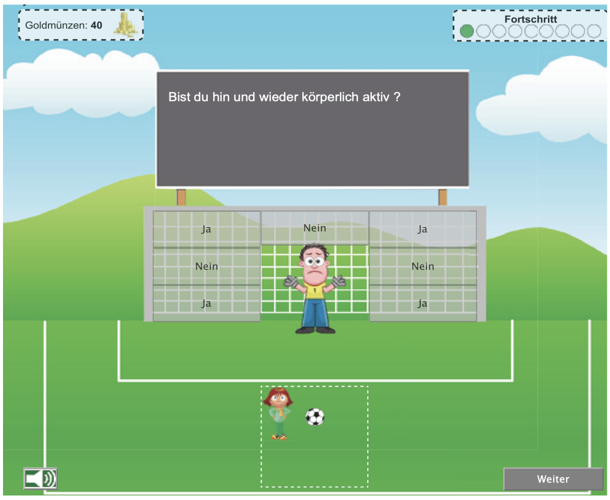


Figure 1. Interface from mini-game-based survey, displaying a closed-ended question. Source: Harms et al. (2015).

Their approach also included other gamification elements such as an internal currency that could be used to purchase game items. Their results showed an increase in participant motivation and the perceived fun, but led to a lower overall response rate and possibly reduced response quality (Harms et al., 2015).

Another study designed a template system for a gamified online survey based on elements from the Bingo game, as displayed in Fig.2. The study reported increased enjoyment and motivation, but without significant differences in response quality (Kuwamura et al., 2021).

Figure 2. Screens from Bingo survey. Source: Kuwamura et al. (2021).

Often, when using gamification techniques a certain degree of increasement in motivation is achieved, but it is important to notice that these strategies can potentially discourage intrinsically motivated participants, thus inadvertently reducing response rates and data quality(Knowles & Brown, 2021). The results from these studies underscore the delicate balance between engagement and efficacy in survey design, as sometimes there is a trade-off. Participants may be more motivated by the overall game experience, but less focused on answering questions thoroughly.

3.2 Data Collection in Citizen Science

Citizen science projects that crowdsource data need to engage the public in data collection processes. In that field, successful citizen science initiatives have leveraged technology, creating devices and applications, to make processes more efficient and comprehensive. Two examples of this, the Citizen Dialogue Kit Project and the Digital Matatus Project, stand out due to their approaches, scalability, and impacts.

3.2.1 The Citizen Dialogue Kit Project

The Citizen Dialogue Kit (2019) is a research project by the Research [x] Design team at KU Leuven, aimed at enabling organizations and initiatives to collect data using interactive public space devices. The devices were designed for polling, and they used e-paper screen for displaying the aggregated data collected, as displayed in Fig. 3.

Figure 3. One of the polling devices in Citizen Dialogue Kit. The device displays the question in the top, the visualization of collected data (graph) in the middle, and has push-buttons in the bottom for each answer. Source: Cornelissen (2021).

These devices are designed for easy deployment, powered by rechargeable batteries, and constructed with mainly off-the-shelf electronics. The kit's methodology includes two workshops to determine topics, audience, data sources, and visualization concepts. The initiative presents a novel loop in data collection, where displayed data serves for co-mmunicating data but also for recruiting new participants, as it encourages passers-by to participate (Coenen et al., 2021; Coenen et al., 2019; Research[x]Design - KU Leuven, 2022).

The project contributes an open-source tool for various contexts, focusing on presenting locally relevant data within its physical and sociocultural context to engage passers-by and spark community dialogue. However, current limitations of it include that the display is rather small, mainly because of e-paper screens costs. Also, the devices only allow for closed-ended questions.

3.2.2 The Digital Matatus Project

The Digital Matatus (2015) project is a citizen science initiative to collect and standardize transit data on matatus, which are privately owned minibuses used as shared taxis in Kenya. Through this project, citizens contributed essential transit data using their mobile phones (see Fig. 4), such as routes, infrastructure available ant the state of matatus, and schedules. The collected data was freely distributed to promote innovation and service improvement (Civic Data Design Lab MIT, n.d.; Digital Matatus, n.d.; Williams et al., 2015).

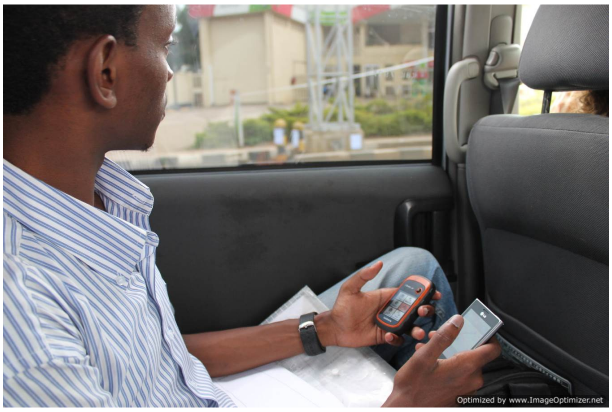


Figure 4. The use of mobile devices for tracking matatus in Nairobi. Source: Digital Matatus (n.d.).

The project's aim is was make public transit more visible, efficient, and open, utilizing technology, local partnerships, and active engagement of the transit community. Crowd-sourced data collection was achieved via mobile apps, resulting in the creation of Nairobi's first transit map and matatus database. Additionally, the project developed scalable data collection tools, with an impact extending to the entire city. Recognized by the City of Nairobi, this data now guides the development of Nairobi's Bus Rapid Transit system. The project underscores the potential of mobile apps for crowdsourcing data collection and demonstrates the importance of stakeholder involvement to achieve the envisioned outcomes.

3.3 Exhibitions as Tools for Data Collection

Interactive exhibitions can be groundbreaking platforms for data collection in citizen science. Despite their smaller scale, they can generate rich qualitative datasets, that if are then digitalized and made accessible, could contribute significantly to research. Although they hardly ever used for this purpose, these exhibitions open up new opportunities for exploring qualitative citizen science research. Two of such are discussed in the following paragraphs.

3.3.1 The Fred Hutch Visitor Center Project

The Fred Hutch Visitor Center project is a compelling example of using exhibitions as a medium for data collection. Commissioned by the Fred Hutchinson Cancer Research Center and brought to life by Studio Matthews in 2017, it offers visitors an opportunity to share their personal experiences with cancer. The centerpiece of this effort is a simple yet effective installation, consisting of a photo booth, made using an iPad, a ring light, and a basic laser printer. Here, visitors can take a photo, print it, write their story, and add it to a growing collection displayed on the walls.

While it is not a citizen science project, the Fred Hutch Visitor Center project provides an illustrative case of qualitative data collection in a public setting. The 'Share Your Story' section showcases hundreds of personal stories and testimonials, giving staff, researchers, and visitors a tangible sense of the impact of the work being done at the Fred Hutchinson Cancer Research Center (Adona, 2017; Studio Matthews, 2022).

(a) A general look of 'share your story'

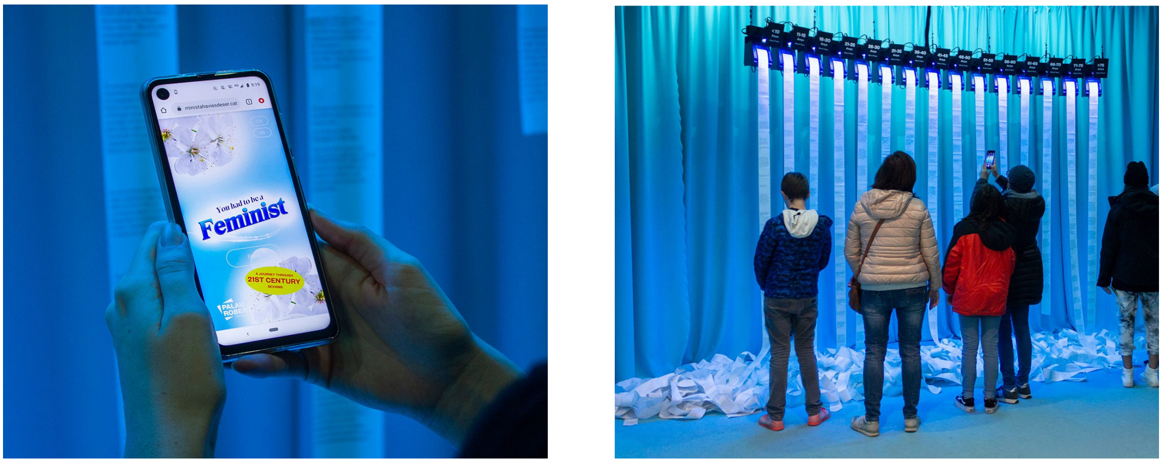
(b) Shared stories

Figure 5. The 'share your story' interactive piece displays collected qualitative data in a format of printed black and white photographs and handwritten text. Source: Studio Matthews (2022).

3.3.2 You Had To Be a Feminist Exhibition

You had to be a Feminist is an exhibition commissioned by the Government of Catalonia, that showcases a novel approach to crowdsource qualitative data on a sensitive topic. Designed and developed by Domestic Data Streamers in cooperation with curator and writer Natza Farré, the exhibition invites participants to anonymously share their experiences of sexism via an interactive piece. A unique aspect of this experience is its COVID-19-conscious design, using a custom website and QR codes for touchless interaction and data submission (see Fig. 6).

Once a participant submits their story, the narrative is printed in real-time and displayed in the exhibition space. Despite the data being for internal use only, the project provides a compelling example of how to collect sensitive qualitative data. This approach to data collection, especially for narrative data around specific issues, could be adopted and adapted by citizen science projects in need of rich, insightful data (Domestic Data Streamers, 2021; Palau Robert, 2020).



(a) The mobile platform (b) Printers with submitted answers

Figure 6. The interactive installation piece in *You had to be a feminist.* Source: (Domestic Data Streamers, 2021).

3.4 Creative Computing

Creative Computing is an emerging area in HCI that uses computation as a medium for creative expression (Yang & Zhang, 2016). This approach seeks to push the boundaries of how we traditionally conceive technology. Various tools can be found in that realm, such as Processing and P5.js, that aim to facilitate the creation of interactive digital experiences.

Processing is an open-source graphical library and integrated development environment (IDE) built for the electronic arts and visual design communities. It aims to teach the basics of computer programming in a visual context (Processing Foundation, n.d.). It offers a simple platform to learn and prototype with and is used by students, artists, designers, and researchers to create a wide range of interactive graphics, animations, and digital art pieces.

P5.js, on the other hand, is a JavaScript library that brings the principles of Processing to the web. It provides an easy-to-use environment to create interactive graphics and animations directly in the browser through code. P5.js enables the blending of web elements, sound, and interaction, thereby broadening the possibilities for creative expression in digital media (p5.js, n.d.).

The possible outcomes are virtually endless. Two very different projects, that share a creative computing approach, are presented in the following paragraph, as a way to illustrate the potential.

3.4.1 The Morphogenetic Creations Project

Andy Lomas's project Morphogenetic Creations is an exemplary demonstration of the application of Creative Computing in digital art. Lomas's creations explore how complex structures, forms, and behaviors can emerge from simple rules and interactions, a process known as morphogenesis. Using custom-written algorithms, his digital art pieces mimic natural growth processes (see Fig. 7), creating intricate and compelling visuals that seem to blur the boundaries between the organic and the artificial (Giannini & Bowen, 2019; Lomas, n.d.).

Morphogenetic Creations showcases the power of creative computing in creating novel aesthetics and pushing the boundaries of digital art. The artist explores the aesthetics of biology, mimicking nature's apparently random systems through algorithms. The project illustrates how computation can be leveraged as a tool for creative exploration and expression, challenging traditional views of technology.

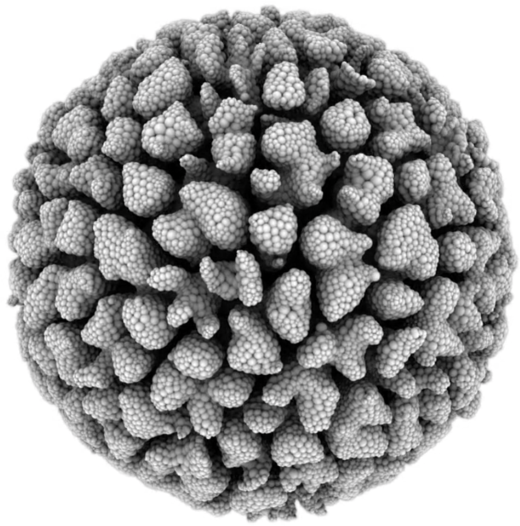


Figure 7. Cellular forms, part of Morphogenetic Creations. Source: Lomas (n.d.).

3.4.2 The Nature of Code project

Daniel Shiffman's The Nature of Code is an innovative project that merges creative computing with education. It's an interactive digital book and videos, that teach programming and the principles of physics, biology, and mathematics through it. Each concept is paired with a set of interactive examples coded in p5.js, allowing readers to modify, experiment, and learn through direct engagement (Shiffman et al., 2012; Shiffman, n.d.).

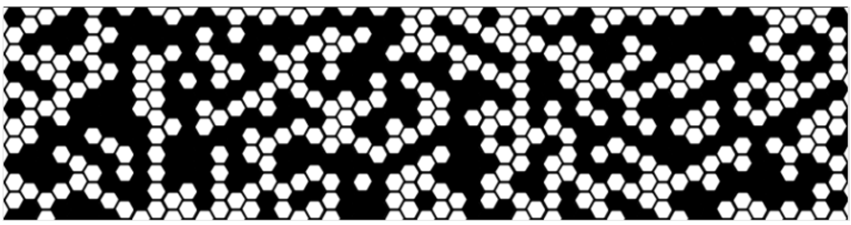


Figure 8. Visual output from a P5.js script on Cellular Automata chapter in The Nature of Code. Source: Shiffman (n.d.).

The Nature of Code exemplifies the transformative potential of creative computing in education. As an educational project, it provides a platform for active learning, enhances the understanding of complex concepts, and makes abstract ideas graspable. The use of p5.js showcases the power of creative computing tools in crafting interactive content, fostering more engaging experiences.

3.5 Analysis

All of the projects examined in this chapter offer distinct methods for crowdsourced data collection, each with its unique attributes and challenges. A summary of them can be found at the end of this chapter, in Table 1. Projects range from online surveys to interactive exhibits and creative software applications. While these approaches differ significantly in their implementation and the type of data they gather, they all share a common aim of engaging the public in contributing meaningful data.

Online surveys, have the ability to collect both qualitative and quantitative data. They are highly accessible and scalable, but their main challenge lies in maintaining engagement and achieving high response rates.

The design of citizen science tools, as highlighted by the Citizen Dialogue Kit and the Digital Matatus Project, significantly impact how easily participants can contribute data and stay involved in the process.

Interactive exhibitions like the Fred Hutch Visitor Center and the You Had To Be a Feminist provide more spontaneous and rich qualitative data contributions. These contributions, often in the form of narratives or personal stories, offer potential for a deeper understanding of individual experiences and societal dynamics. Even without elaborate visualization or aggregated data displays, the collected information arouses interest as visitors navigate between stories and explore different data entries.

Creative Computing projects, while varied in nature, as exemplified by Morphogenetic Creations and The Nature of Code, can also offer means for interactivity, and engagement. Even though they are not yet used for data collection, they can be leveraged to bring interactivity into its processes.

Also, the combination of physical and digital tools, as seen in the exhibition created by Domestic Data Streamers and in the devices from Citizen Dialogue Kit, presents an interesting approach, exemplifying how these two types of tools can be interconnected for a more holistic experience.

In conclusion, these projects demonstrate that the success of crowdsourcing data collection depends not only on the act of collecting data, but also on the design of tools that facilitate meaningful and engaging interactions. They underscore the need for a balanced approach that combines ease-of-use and accessibility with the ability to capture rich, insightful data. The nuanced analysis of these projects illuminates the opportunities and challenges in the field, providing a foundation for future exploration and innovation.

3.6 The Opportunity for Creative Computing in Qualitative Data Collection for Citizen Science Projects

The projects presented illuminate the potential of mindful design for voluntary qualitative data collection in citizen science initiatives. Creative computing presents a rich array of strategies, offering unique entry points throughout the lifecycle of a project.

Visualizing this lifecycle as a timeline, it begins with data collection, where participants can voluntarily contribute entries. In this stage, creative computing, at the intersection with HCI, can transform the ways in which people participate and contribute information. Creative computing strategies could potentially simplify the process, make it more engaging, enhance participation, and increase data quality.

When a project requires ongoing data entries from participants, creative computing can further play a crucial role. This stage calls for maintaining participant engagement. The intersection of data visualization and generative art could produce powerful visual representations of participants' contributions. Providing an interactive and visually engaging approach , participants could potentially better understand the immpact of their contributions and overall significance of the project.

Finally, the culmination of the citizen science project involves analyzing the collected data to extract insights. These insights often influence decision-making among key stakeholders, but importantly, should also be fed back into the community. Creative computing also presents an opportunity at this stage, in the way these analyses and insights are communicated. It can help meet different levels of understanding among the audience. Also, it offers the potential to make the insights not only relevant to the deeply involved participants but also to those encountering the project for the first time.

Therefore, the potential of creative computing approaches is relevant throughout the lifecycle of citizen science initiatives, providing an array of strategies to enhance participant experience and engagement, ultimately aiding in better data collection, understanding, and dissemination.

Project or study name	**Category**	**Summary**	**Highlights**
Messenger survey	Online survey	Messenger-app design for survey	Longer completion time, shorter answers
Mini-game based survey	Online Survey	Mini-games applied to survey	Challenging to implement. Increased motivation, decreased response quality
Bingo survey	Online survey	Survey that imitates Bingo game	Increased motivation, no significant impacts on data quality
Citizen Dialogue Kit	Citizen science project	Public polling devices with data displays	Novel data collection device. Only allows for closed-ended questions
Digital matatus	Citizen science project	Mobile apps for crowdsourcing transit data (matatus)	Successful project, impacts beneficial for the community
Fred Hutch Visitor Center	Exhibition	Participatory installation gathering personal narratives	Very rich qualitative data volunteered. It is not systematized nor used after
You had to be a feminist	Exhibition	Participatory piece in exhibition gathering personal narratives	Rich text narratives on a complex issue. Volunteered data
Morphogenetics	Creative Computing	Generative art mimicking biology	Illustrates the potential of generative algorithms' aesthetic value
The nature of code	Creative Computing	Book and videos that teach programming and the principles of physics, biology and mathematics through code	Demonstrates the potential of creative computing for teaching but also for online engagement and interactivity

Table 1. Overview of the projects and studies presented

4 Methodology

The objective of this research is to investigate how creative computing techniques can enhance the data collection process in online qualitative surveys, enabling the collection of rich and insightful data. This study proposes the concept of generative data representations, which are introduced and described in this chapter. They are conceived as a way to engage participants responding open questions in online surveys. To validate this hypothesis, experiments were conducted to compare answers from a traditional survey design (non-canvas), with a survey design that incorporates the generative data representations (canvas). The objective was then to investigate the impact on engagement of participants responding to open questions with long text inputs types.

This chapter provides an overview of the research methods and techniques employed to conduct the study. It presents the research approach selected for this study, introduces the concept of generative data representations, and with them, the different data dimensions that are encoded into features. Finally, a diagram on how the methodology and experiments are interconnected is presented, to delve deeper into them in the next chapter.

4.1 Research Approach

To assess the contribution of creative computing, an experimental research design approach was employed. This approach offers greater control over the factors being studied, allowing for more conclusive evidence of cause-and-effect relationships (Dubey & Kothari, 2022). By conducting experiments, it is possible to determine whether the proposed solution aligns with the hypothesized outcomes by comparing the effects of specific variables. This approach was chosen because the aim is to generate more generalized knowledge. The experiments allowed for original data collection, and mixed methods were used to gather qualitative and quantitative data. A quantitative approach was used for variables that are more objective and can be quantified (Dubey & Kothari, 2022), while a qualitative approach was used to complement with opinions, preferences and participant's behaviors, and thus get deeper insights.

4.1.1 Creative Computing

As mentioned in previous sections, intrinsic motivators are considered more suitable for engaging participants in online surveys. Meaningful interactions can act as intrinsic motivators, thus highlighting the role of Interaction Design. When used mindfully, IxD can create the sense that what people are saying is valuable for the study. In other words, as if they're being *actively* listened to, mimicking mediated interactions present in face-to-face surveys. Therefore, this study uses creative computing techniques to enhance interactivity and engagement, and elicit deeper insights from participants in their responses.

4.1.2 Data Representations for Human-Computer Interactions

It is important to clarify that this study is designed within the realm of Human-Computer Interaction (HCI), rather than being a data visualization project, as the objective is to create an engaging and participatory experience for respondents. Data visualizations often focus on encoding data visually to facilitate its understanding and interpretation (Cairo, 2016). This study draws inspiration from data visualization, but mantaining the aforementioned objective.

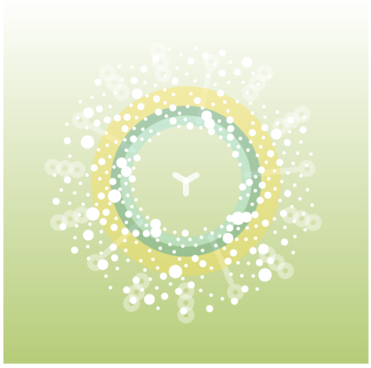
In the approach proposed in this study, participants actively contribute to the creation of visual representations while participating in the survey. This interactive approach also draws inspiration from interactive installations, where individuals become co-creators, in this case, in the process of generating visual outputs. By involving participants in the visualization process, the study aims to motivate people to be more detailed in their responses, and foster a sense of ownership over the generated representations. To achieve this, generative data representations are introduced.

4.2 Generative Data Representations

In this study, the term generative data representations refers to the visual outcomes generated through an algorithmic encoding of data. Specifically, the data takes the form of free-text responses obtained from open-ended questions in online qualitative surveys. During the data collection process, participants' text inputs are analyzed in real-time, resulting in dynamic interactions where the representations are created as they respond. Figure 9 displays how data from free-text gets encoded into a visual output.

"University has affected my life in a positive way, since I am motivated to be able to work in what I study and proud to face challenges, it has helped me a lot to get out of my comfort zone, to enjoy socializing, and to be able to cope with things that previously affected me, also my institution provides me with a free psychologist who, along with my friends, has helped me a lot too"

(a) Text answer



(b) Answer representation

Figure 9. Generative Data Representations. In 9a the raw data is displayed, while in 9b the same data is displayed as a visual representation.

Although participants may not be aware of the specific expected outcomes of the generative process in terms of which action generates which reaction, it does not obstruct the objective of eliciting interactivity. Performing the analysis of text in real-time is what sets apart this approach from other methods of data visualization.

Generative data representations 'react' to what is being typed, and for participants, it is that what creates interactivity, as data representations are constantly adapting to what the participant is typing, changing its visual configuration to represent the content. The changes triggered by some specific words or phrases draw a parallel to the nods and non-verbal cues that a mediator may have in face-to-face interactions, as an indicator to detail more when a topic that emerged in the answer is relevant.

4.2.1 Speculating Answers

Traditionally, researchers encode data *a posteriori* meaning that all the data is first collected and then analyzed. In this way, it is possible to determine the dimensions of the representation and generate visual elements, ranges, and configurations. The proposed approach presents the additional challenge of creating these encodings *a priori*, as the data will only be available in later stages, after participants have answered the questions. Therefore, instead of designing data visualizations, it is necessary to define structures that will be completed with the answer's content. A certain degree of speculation is required for this approach.

To draw an analogy, if it was working with quantitative data this approach would be similar to creating a visual parametric design, in which parameters are completed with the values of the inputted numbers. In this study, instead of numbers, the system works with free-text that is multidimensional, and has varying degrees of data richness and lengths. Working with unstructured data, with such variable structures, represents an important challenge. Because of that, it was preferred to work with a keyword-detecting approach. Three categories of keywords (described later in this section) were defined, to provide some degree of structure.

The proposed system can be customized to address the different topics and objectives that a citizen science qualitative research may have. Determining the data dimensions, as well as the data type of the expected answers is directly linked to the research objectives of that research. Specific keywords and topics that are important to explore in more depth can be defined through analysis, revising literature, and other research methods.

4.2.2 Data Dimensions

The MID system (Hil & Lachenmeier, 2022) defines five different data types: location, text, time, category, and figure (numbers). Each data type can be used to form categories, based on the research topics and keywords selected. In this research, data in the form of free-text is grouped into three categories or dimensions: emotions, entities, and actions. The reasoning behind is that for researching people's perceptions and experiences, these three dimensions provide greater differences in content analysis.

Emotions are color-coded using the Emotion Typology, a systematic classification of emotions according to their differences and similarities developed by the Delft Institute of Positive Design (of Positive Design, 2016). Their research provides formal definitions and additional resources for further reading. It has contributions from researchers across multiple universities. The list of emotions, with their respective colors, can be seen in Fig. 10. The other two categories defined, entities and actions, are encoded using shapes. The logic behind this is explaining in the following subsection.

4.2.3 From Dimensions to Features

Representations are composed of various features that together create multi-dimensional encodings. As mentioned, there are three main data dimensions: emotions, entities

Figure 10. Positive and negative emotions color coded, from Emotion Typology. Source: of Positive Design (2016).

(subjects), and activities. Data dimensions get encoded by features, and the sum of all features composes the generative data representation. An overview of this approach can be seen in Fig. 11.

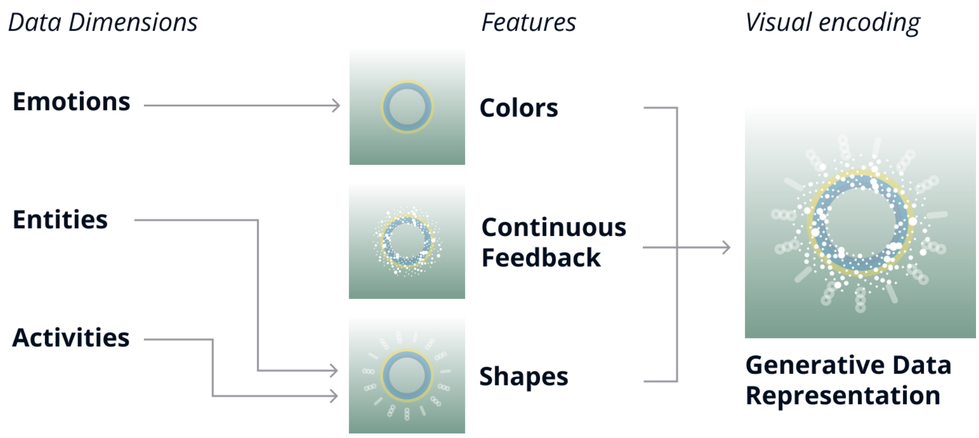


Figure 11. Data dimensions get encoded into features. The generative data representation displays the combination of all features

Additionally, a continuous feedback loop was implemented, that slightly modifies the output responding to each key pressed by the respondent. The platform features a dynamic background that displays the respondent's mentioned emotions, while different shapes and their distribution in the canvas represent other relevant keywords and topics. Each feature is explained later on this chapter.

4.2.4 Visual Complexity

Encoding the dimensions through the aforementioned features allows for varying degrees of visual complexity. A group of different answers representations can be seen in Fig. 12.

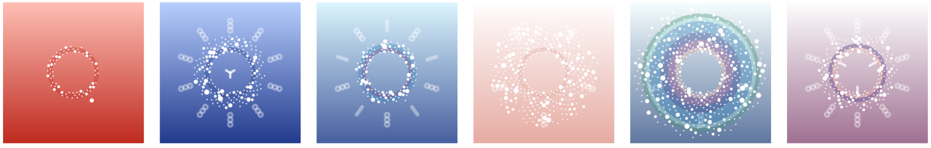


Figure 12. Examples of different visualizations from six different answers.

The complexity of the representation is closely tied to the richness of the input data. Specifically, the more keywords and relevant information are included in respondents' inputs, the greater the number and diversity of elements that are displayed in the canvas. Fig. 13 provides a side-by-side comparison of two answers with similar word counts, but differing levels of richness. This comparison illustrates how variations in input quality can have a significant impact on the resulting representation.

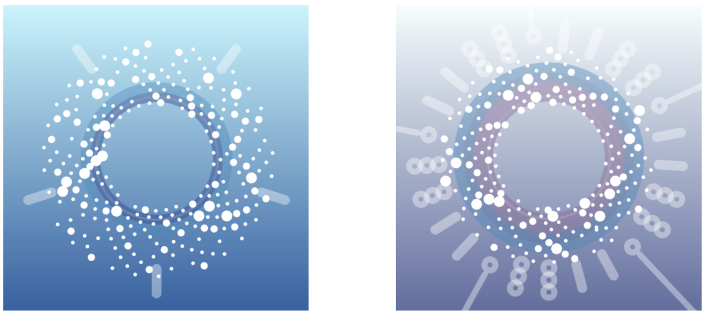


Figure 13. Representations with similar word counts. The one on the left has fewer keywords than the one on the right.

4.3 How Representations are Generated: Features

The proposed system involves many variables, and so far it has only been addressed in a general way. In this section, each feature is better explained to fully understand how generative data representations are created.

4.3.1 Continuous Feedback Loop

The idea that a visual representation can adapt to keywords and specific topics, is not something respondents would be familiar with. In the first interaction, it may be difficult for them to understand how the system works, so the continuous feedback interaction was designed. Each pressed key generates a dot on the screen so that each action modifies (subtly but concisely) the visualization. The dots correspond to the letters typed, different letters have different dot sizes based on their frequency of usage, and characters that are not letters are rendered transparent. The configuration of compositions can be seen in Fig. 14.

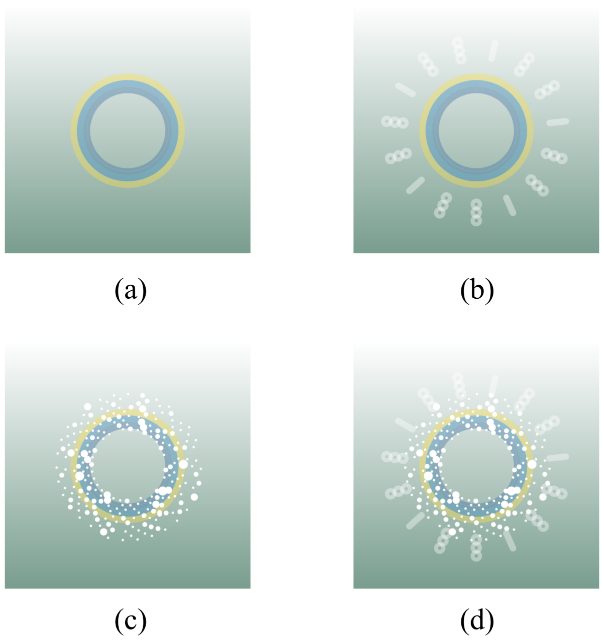


Figure 14. Different representation features. Fig. 12a displays only the background gradient and a color ring for every emotion detected. Fig. 12b displays the background gradient with the continuous feedback loop, while Fig. 12c does it with shapes instead. Finally, Fig. 12d shows the full generative representation.

4.3.2 Background Gradient

As mentioned, each emotion has its own color code associated, which is used to create a gradient in the background of the canvas. The use of these color codes makes it possible to work with emotions as numerical variables (RGB), allowing for an aggregated visualization. If respondents mention more than one emotion, the codes generate an aggregated color. Each code has 3 dimensions or channels: red, green and blue. They are averaged individually and result in another aggregated color. In this way, the background displays the combination of all the emotions mentioned, factoring also the number of times it is mentioned. This component generates a notorious visual change in representations as displayed in Fig. 14. Additionally, each emotion is displayed individually as color rings in the center

4.3.3 Shapes

The remaining categories, namely entities and activities, can be organized in two dimensions: *self* and *environment*. Both are encoded through shapes. The *self* dimension corresponds directly to the respondents themselves. On the other hand, the *environment* dimension relates to their connections, interactions, and systems in which they are involved, including the entities and activities around them. The assignment of specific words or themes to each of these categories is context-dependent, as it is influenced by the research topic and scope.

In the center of the visualization, there is a circle that acts as the boundary between *self* and *environment* dimensions. Keywords and themes related to the participant are represented inside the boundary towards the center, and those related to the environment are represented outside (see Fig. 14b).

4.4 Research Design

To assess the impact of the aforementioned approach on the data collection process in online qualitative surveys, it is essential to consider both quantitative and qualitative perspectives. From a quantitative standpoint, it is important to evaluate whether the inclusion of generative data representations enables researchers to obtain richer answers by contrasting different numerical variables. Since there are many factors affecting that, such as survey design, the research topic itself, the questionnaire length, and how questions are configured (Galesic & Bosnjak, 2009), it is important to also assess the user experience to gather insights of what they interpret from the survey and how are they interacting with it. By examining both quantitative and qualitative aspects, it is possible to gain a comprehensive understanding of how the approach influences data collection in online qualitative surveys.

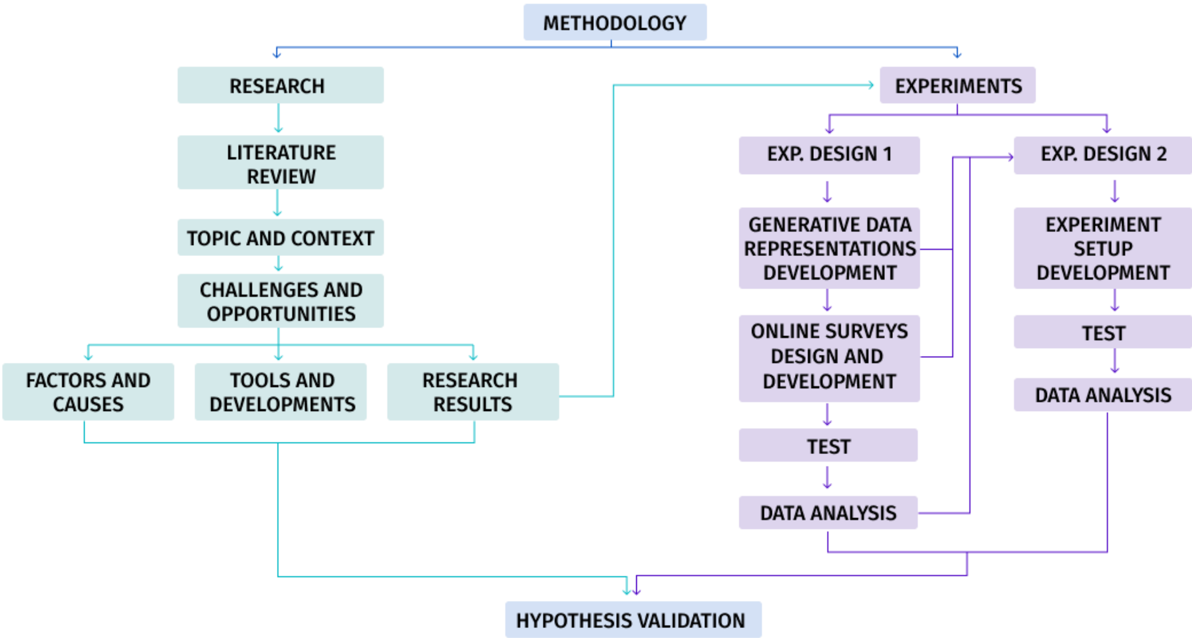


Figure 15. Overview of the proposed methodology for the study.

Based on this distinction, a set of two experiments were conducted. An overview of the research methodology is presented in Figure 15. The next chapter delves deeper both experiments, as each of them is detailed in terms of its design, recruitment process, variables examined, and tools utilized.

5 Experiments

So far the methodology was only discussed on a general level, but inn order to assess the effectiveness of the proposed approach, it needs to be applied into an use case. In these first subsections, the use case is detailed as well as the developed interfaces for it. The subsequent sections of this chapter explain in detail each experiment design and implementation, data analysis approach, and ethical considerations. A summary of the experiments is presented in Table 2.

	Experiment 1	**Experiment 2**
Experiment Design Overview	Contrasting answers with and without generative data representations	Semi-structured interviews after a survey with generative data representations
Recruiting	Non-probability sampling. On-campus recruiting, QR codes	Convenience sampling. On-campus recruiting
Variables	Character count, keyword count, and number of suggestions in responses across the two survey types	Participants' perceptions, interaction patterns, and direct feedback
Tools Used	Custom-made survey	The same survey, a semi-structured questionnaire for interviews, audio recording
Time	One month	Two weeks
Participants Recruited	106 participants total	37 participants

Table 2. Overview of the experiments design and implementation

5.1 Use Case: Mental Health in University Students

The presented approach has so far been discussed in a general sense, as the system has many potential applications across various fields that can benefit from qualitative data (i.e., market research, public opinion polling, healthcare, among others). However, to fully evaluate the effectiveness of the approach, it is essential to provide an illustrative use case within a specific context.

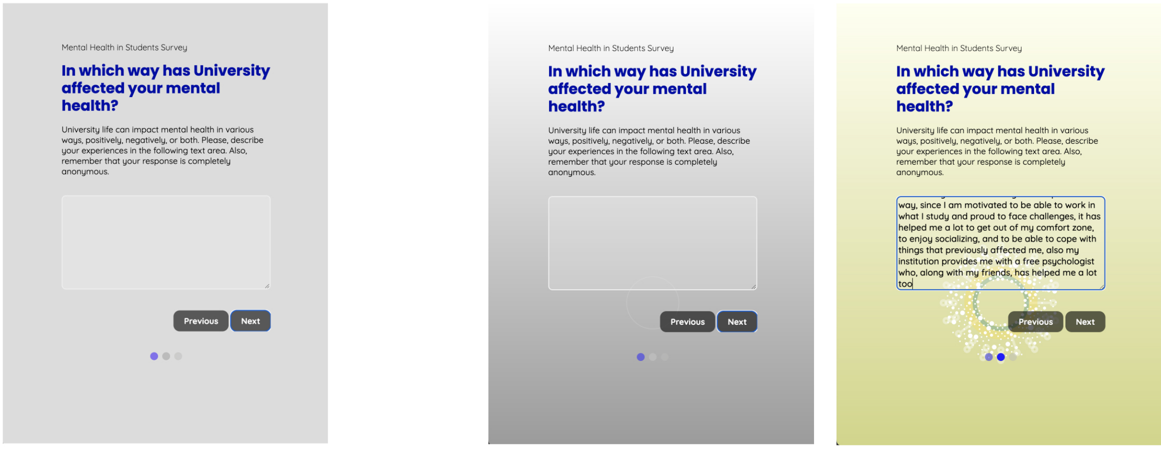
In this study, a fictional citizen science project focusing on improving the mental health of university students was employed as an illustrative use case. In the initial stages of a citizen science project like this, an online survey can be utilized to identify cross-cutting themes that require further investigation in subsequent stages. For this study, that first exploratory survey was selected to test if the use of generative data representations to gather richer and more comprehensive data.

Having investigated the main causes of mental health issues in university students through a literature review, the corresponding keywords and sub-topics were generated. Regarding the *self* category: positive actions (i.e., getting involved, asking for help, self-care), negative actions (i.e., alcohol and drug use, lack of sleep), characteristics (i.e., gender, ethnicity, disabilities). Regarding the *environment* category: entities (i.e., professors, peers, family), positive actions (i.e., support, inclusiveness, socializing), negative actions (i.e., harassment, bullying, discrimination), contextual characteristics (i.e., socio-economic status, workload).

5.2 Interface: Online Survey

In order to implement and evaluate the approach, it was essential to design a suitable platform for participants to provide their responses, so an online survey interface was created. Initially, there was a consideration of contrasting a custom-made platform with a standard form such as Google Forms. However, as the literature review shows, there are various factors, including survey design, questionnaire length, and configuration, that could potentially influence the participants' answers. Thus, it was decided to use the same design for layout the custom-made platform with the proposed system and the standard form.

The developed interface maintains the structure of a standard form, where questions are displayed along with a text field to type answers and a button to submit. This study introduces a canvas (which holds the generative data representations) in the background of the form, a novel element that allows displaying and modifying the representations of the answer in real-time.



(a) non-canvas (b) canvas empty (c) canvas filled

Figure 16. Form interface. Standard form with no canvas is displayed in Fig. 16a, while Fig. 16b and Fig. 16c correspond to the interface with canvas in the background, empty and with an answer respectively.

The system was designed to prioritize mobile compatibility, while still functioning similarly on desktop devices. In the layout design, the canvas was placed directly behind the input element for improved usability. This is because if placed on top, when the keyboard was activated, all elements would be shifted upward and may be out of view. For this study it was essential that respondents had an unobstructed view of the canvas without needing to scroll. The interface layout for both forms can be see in Fig. 16.

As shown in Figure 15, the experiments were designed based on the literature review. Each experiment is detailed in this section, in terms of methods, implementation, tools and variables.

5.3 Experiment 1: Quantitative Approach

The objective of this first experiment was to determine whether visually representing a long text input response in real-time produces richer data. It involved contrasting the answers to open questions collected through questionnaires, one with generative data representations and the other without, to see if there was an effect in the outcomes. The experiment is described in terms of experiment design, recruiting, variables, tools, and time in the next paragraphs.

5.3.1 Experiment Design

The design of this experiment allowed for contrasting the results obtained through two open-ended qualitative survets. A traditional online form was opposed to a custom-made platform that generates real-time visual representations of what a participant is typing. The expected result of the design of this experiment was to have a platform that creates and modifies an abstract vector image with each key pressed by the user, and that as a whole, represents the response received. In this way, it was also expected that the representations obtained from different responses would be diverse and distinguishable.

Recruiting: As participants needed to have similar characteristics, working with students ensured a certain degree of uniformity, and to address potential sampling biases, additional questions were included in the survey to determine the participants' university and study program. It was important to balance the number of responses between the canvas and no-canvas forms, so a system was implemented to automatically redirect participants to the form with fewer responses in the database. The recruitment process involved handling and displaying QR codes on campus to invite university students to participate. The objectives of the study were not disclosed to avoid biases.

Variables: The selected variables aimed to capture the characteristics of the obtained responses, enabling comparisons on data richness. Although the platform design differed slightly between the canvas and no-canvas surveys, the questions themselves were identical. The quantitative measurements focused on comparing character counts, keyword counts, and the number of suggestions provided by participants.

Tools used: A custom-made survey was developed. It used a node.js server connected to a database to store the answers, and dynamically determine which version of the survey was presented. To disseminate the survey QR codes where made that linked to the survey consent page. For the creation of the form, generative graphics were created using p5.js library, while the rest of the platform used HTML, CSS, JavaScript and Node.

Time: Both surveys were made available simultaneously, but each person could only see one version of the survey. Overall, experiment took about a month to deploy, because minor changes in the format or improvements required to start over. A pre-testing period consisted on trying out different experiment settings, assessing how they performed and adjusting it accordingly. Also, it was important not to recruit people that had already answered a variation of the experiments to avoid biases.

5.3.2 Implementation

The survey consisted of three sections presented on separate pages, employing a multi-step format. In the first step (Q0), participants were asked to provide their program of studies and institution in two short input areas. This step aimed to facilitate the identification of potential biases, such as outliers from the same institution or program.

The second step involved an open-ended question (Q1) that asked participants, *'In which way has University affected your mental health?'* accompanied by the caption, *'University life can impact mental health in various ways, positively, negatively, or both. Please, describe your experiences in the following text area. Remember that your response is completely anonymous'.* The input area for their response was a large text area with 8 rows, and it was presented in both the canvas and no-canvas versions.

The third and final step presented another open-ended question (Q2) that inquired, *'How can your university better support students' well-being?'* alongside the caption, *'Please share your thoughts on what specific actions or resources your university could provide to better support students' mental health and overall well-being'.* Similarly to Q1, this question was accompanied by a large text input area and was available in both the canvas and no-canvas versions.

The logic behind the question sequence was that Q1 was highly personal, which could make it more challenging for people to answer, potentially resulting in shorter responses. On the other hand, it could also be more relevant to participants, leading to longer answers. Therefore, the second question aimed to be less personal and more practical in nature.

A total of 107 participants, comprising undergraduate and graduate students from various disciplines, were recruited on campus through QR codes on posters, projections, or printed materials. The sample was diverse, encompassing individuals from a wide range of backgrounds and experiences. Out of the total group, 53 participants completed the canvas version, while the remaining 53 participants completed the non-canvas version of the survey.

5.4 Experiment 2: Qualitative Approach

The objective of the second experiment was to gain qualitative insights into how individuals interact with the system. To achieve this, the canvas survey was administered through in-person interviews. Participants were provided with the opportunity to respond to the survey questions in their own devices, and subsequent discussions were conducted to capture their insights on a predefined set of topics.

5.4.1 Experiment Design

For this experiment, a semi-structured interview was designed to gather qualitative insights. Participants were provided with a link to respond the canvas version of the survey. While they were responding, participants were encouraged to provide comments aloud but were not allowed to ask questions to avoid potential biases. To ensure privacy and simulate their typical response behavior, participants used their own devices. After completing the survey, a set of three topics were discussed to capture further insights and perspectives.

Recruiting: participant recruitment process involved a convenience sampling method conducted on a university campus. Potential participants were approached individually, and their participation in the study was requested. This approach aimed to ensure diversity and minimize selection biases.

Variables: The variables examined in this experiment included participants' perceptions (i.e., insights into how participants perceived and interpreted the generative data representations), this involved analyzing their subjective impressions, understanding of the visualizations, and any potential connections they made between the visual elements and their responses. Also interaction patterns, or how they navigated through the survey and engaged with the generative data representations, and direct feedback (i.e., capturing their thoughts, impressions, suggestions, or any other comments related to their experience and the overall process).

Tools used: For Experiment 2, a version of the same platform used in Experiment 1 was deployed, with the distinction that it consistently linked to the canvas version. This was done to ensure that participants interacted with generative data representations throughout the survey. In addition, audio recording was utilized during the discussion phase to facilitate conversational analysis. Participants were asked for specific consent regarding the recording before proceeding.

Time: The interviews for Experiment 2 were conducted over a period of two weeks. This duration allowed for a sufficient number of participants to be recruited and interviewed while ensuring efficient data collection. The time frame also accounted for any necessary adjustments or rescheduling that might have been required during the interview process.

5.4.2 Implementation

For Experiment 2, participants were provided with a link to the survey and were instructed to complete it using their own devices, and given space to write their answers privately. Throughout the survey, participants had the option to make comments aloud, although they were explicitly instructed not to ask questions during the survey process. This was to mimic the conditions of Experiment 1, which was asynchronous, thus not allowing for real-time inquiries.

After completing the survey, each participant participated in a semi-structured interview to delve deeper into their experiences and perspectives. The primary objective of the interviews was to gain insights into participants' overall thoughts on the survey and their experience in completing it. Participants were encouraged to share their perspectives on the generative data representations and discuss the possible meanings associated with those representations. Furthermore, participants were asked whether these representations had any influence on their answers. The semi-structured nature of the interviews allowed for flexibility in exploring additional relevant topics or probing for further details as needed.

A total of 37 participants took part in the study, with recruitment conducted using a convenience sampling method. Students were approached in person and directly requested to participate in the study. The participants consisted of university students from various disciplines and institutions.

To ensure that all participants had a first-time experience with the proposed system, measures were taken to ensure that none of the participants had previously taken the survey. This step aimed to minimize any potential bias or familiarity with the system, allowing for fresh perspectives and unbiased feedback.

5.5 Data analysis and Ethical Considerations

The obtained data from both experiments consists of qualitative and quantitative data. Regarding the quantitative data, statistical analyses were conducted, taking into account the size of the data set. On the other hand, the qualitative data analysis focused on the content of user responses in the interviews. Thematic analysis was applied to identify common topics and interesting insights. This analysis involved grouping responses based on thematic similarities, while also accounting for borderline cases, enabling a more detailed examination of the experiment's results.

Ethical considerations were an important aspect of this study. Some ethical considerations that were relevant included:

Informed consent: Obtaining informed consent from participants was important to ensure that they understood the nature of the research and were willing to participate. This included information about the purpose of the research, and the participant's right to withdraw from the study at any time. This was applied to both the online survey (Experiment 1) and interviews (Experiment 2).

Confidentiality and anonymity: Protecting the confidentiality and anonymity of participants was important to ensure that they feel safe and comfortable sharing personal information. The data collected did not require any personal information to ensure participant anonymity.

Privacy: Respecting the privacy of participants was important to ensure that they felt comfortable and safe during the research process, specially in Experiment 2. Participants that were recorded were be asked for consent before, ensuring an explicit authorization.

Fairness and equity: Research was conducted in a fair and equitable manner and did not discriminate against any particular group or individual based on characteristics such as race, gender, age, sexual orientation, or socioeconomic status.

6 Results

In this chapter, the results of both experiments are presented and discussed, offering insights into the effectiveness of the approach presented throughout this study. The purpose of it is to present the data collected and observations, in a structured manner, to serve as a foundation for a comprehensive exploration of the implications and significance of the findings. This chapter also presents a discussion section that elaborates on the interpretation of the results, considering their implications within the broader context of qualitative research in citizen science.

6.1 Experiment 1: Quantitative Results

The first experiment aimed to determine whether visually representing a long text input response in real-time would result in richer data. A custom-made survey was developed, contrasting traditional open-ended qualitative questions with a custom platform that generated real-time visual representations of participant responses. This subsection presents the quantitative findings obtained from Experiment 1, providing insight into the effects of the generative data representations on participant engagement and the characteristics of the collected data.

6.1.1 Quantitative Analysis Approach

In this study, several variables were examined to quantitatively analyze the responses obtained from the participants. The variables considered in the analysis included word count, character count, and detected keywords for Q1 and Q2. Additionally, the number of suggestions provided in Q2 was also considered.

To calculate the character count, the length function in Excel was utilized. A small program was developed for word count to split the response string and count the individual words. Upon reviewing the initial results, it was observed that the word count and character count analyses yielded similar outcomes. After evaluation, the character count was determined to be a more effective metric as it offered a finer level of granularity. Therefore, the analysis of results only considers character count as a measure of answer length.

For the analysis of detected keywords, the same approach and definitions as the system responsible for generating the generative data representations was utilized, after they were identified by the system, a counter of all keywords detected was set.

Regarding the number of suggestions, a one-to-one analysis was conducted to examine the suggestions provided by participants. Additionally, an AI-based approach was employed to verify if the AI system would output the same number of suggestions for borderline cases, ensuring consistency and accuracy.

The statistical analysis primarily involved calculating the average of the variables under investigation along with analyzing minimum and maximal values. Initially, alternative statistical methods such as a trimmed mean of 10 percent and 20 percent were explored to assess their impact on the results. However, the findings from these variations were not significantly different to the overall sample. Upon consideration, it was deemed appropriate to report the results based on the entire dataset.

6.1.2 Response Analysis

Overall, participants demonstrated a high level of engagement and provided substantial responses throughout the survey. On average, the responses for Q1 had an average length of 217 characters, while the responses for Q2 averaged 163 characters. This may be due to the nature of the questions, where the former is more personal and the latter more practical.

The length of the responses varied significantly, with the lowest response being just nine characters long and appearing in the Q2 non-canvas version. Intriguingly, this succinct response comprised one suggestion. This highlights the fact that length as a variable does not completely capture the richness of the answer. This highlights the importance to include the quantity of recommendations in the quantitative analysis.

On the other hand, the longest answer observed was 726 characters, which occurred in the non-canvas version of Q1. This lengthy response contained six detected keywords, exceeding the average number of keywords per response.

When considering the combined character count (the sum of characters in Q1 and Q2), the minimum recorded was 50 characters. From these, 28 characters corresponded to Q1, and 22 characters to Q2, no keywords were detected in Q1, while one was detected in Q2, and one suggestion was provided. This response was obtained from a participant who completed the non-canvas version.

On the other end of the spectrum, the maximum combined character count reached 1250 characters. This consisted of 686 characters in Q1, 564 characters in Q2, six detected keywords in Q1, four detected keywords in Q2, and three suggestions. This response was obtained from a participant who completed the canvas version of the survey.

It is important to note that conclusions regarding the response rate cannot be drawn due to the way the system was designed and implemented. Data was saved only for participants who completed all three steps of the survey, and incomplete surveys were not saved for analysis. Therefore there are no indicators on each response rate or completion rate. The system was designed to automatically redirect users to complete the version with fewer responses. This approach aimed to ensure an equivalent number of responses for comparative analysis between the two versions at a later stage.

6.1.3 Number of Characters Analysis

The analysis of the responses revealed that, on average, Q1 generated longer responses than Q2. In total, 23,259 characters were collected in Q1, whereas Q2 yielded 17,436 characters, 5,823 characters fewer than Q1.

This outcome is consistent when the nature of the questions is considered (Q1 being highly personal and Q2 more straightforward). As mentioned, the average number of characters per response was 217 for Q1 and 163 for Q2. When combined, the average number of characters for both responses reached 380.

In the non-canvas version, the combined average character count was 366 (213 in Q1 and 153 in Q2), while in the canvas version, it increased to 395 (222 in Q1 and 173 in Q2). This indicates an average increase of 29 characters for the canvas version, with a specific average increase of 9 characters in Q1 and 20 characters in Q2.

Figures 17 and 18 present graphs comparing the number of characters obtained in each response, with a distinction between the canvas and non-canvas versions. As depicted in the graphs from Figure 17, the tendency to have a higher number of characters in the canvas version becomes more apparent in responses with fewer characters than the overall average. From Fig. 18, it can be extracted that the increases on character counts in the canvas version are consistent in both questions, but slightly bigger in Q2.

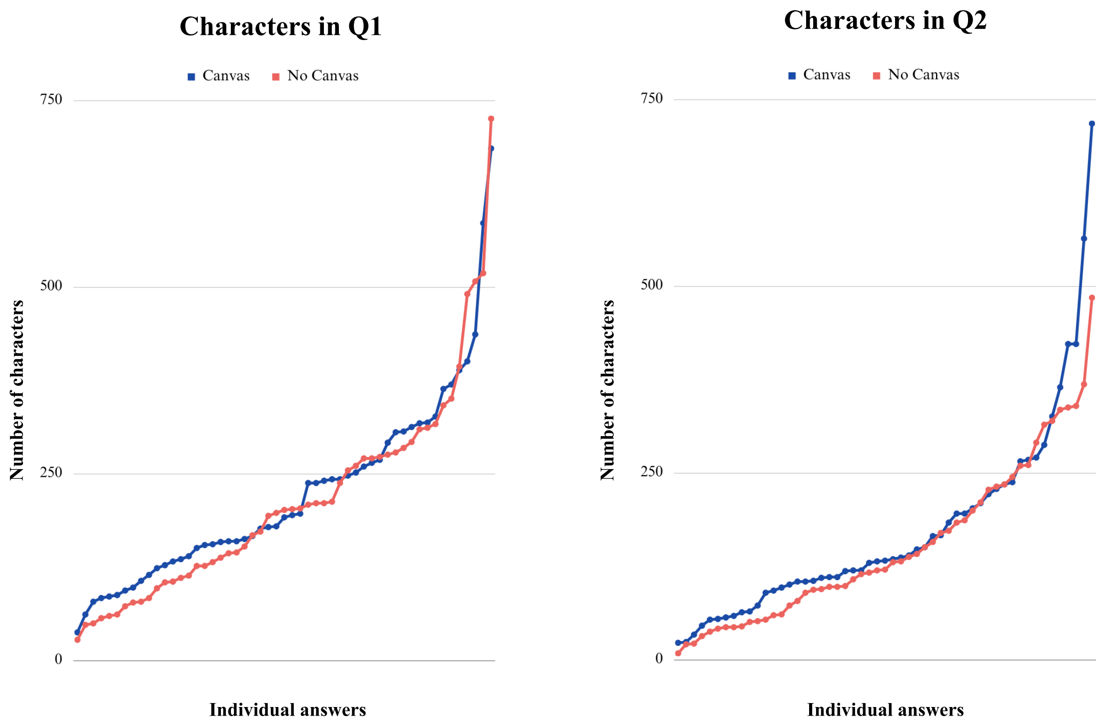


Figure 17. Number of characters by answers, separated by survey type (canvas/non-canvas). The Y axis represents the number of characters detected for each answer, in the X axis all answers are presented and arranged in ascending form. The graph on the left side is from Q1 answers and on the right side from Q2 answers.

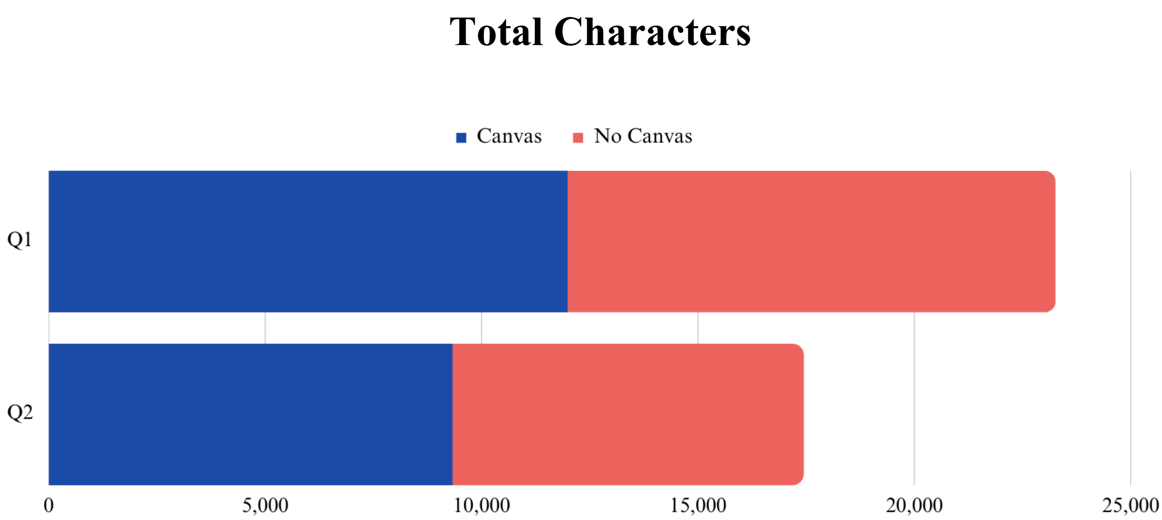


Figure 18. Total of characters yielded from both Q1 and Q2 in the canvas and non-canvas versions.

6.1.4 Keyword Analysis

As previously stated, keywords included in the analysis are the same ones defined in the system specifically for this use case. These keywords are responsible for generating visual changes in the system in the canvas version of it. The number of keywords plays a crucial role in the richness of the responses and the participants' experience. Participants who provided responses without any keywords only experienced the continuous feedback feature without any shape or color changes. Of course, participants in the no canvas version did not experience any changes in the interface, regardless of the keywords count.

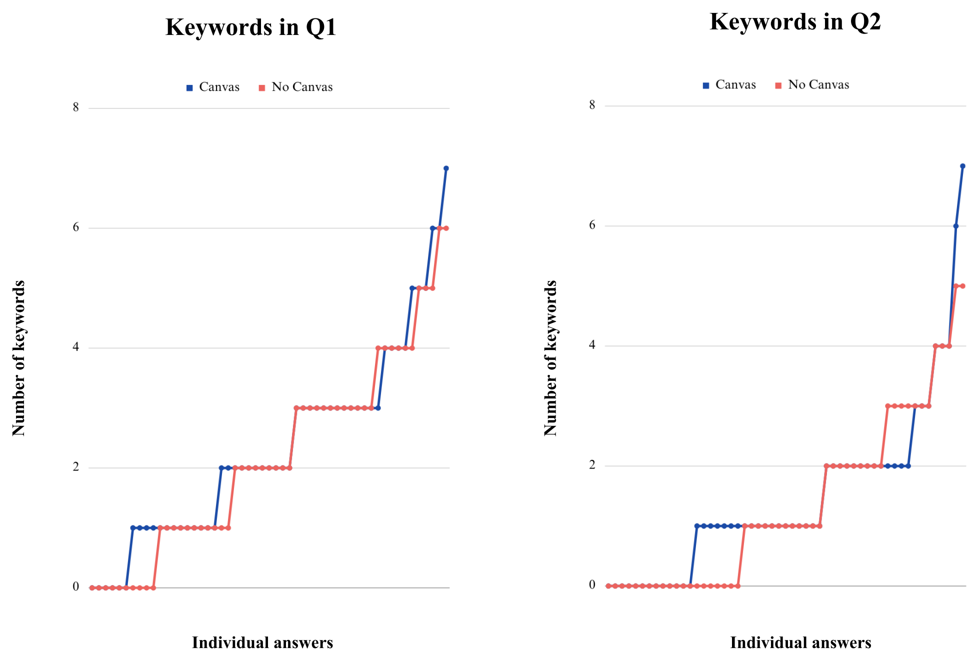


Figure 19. Comparison of keyword counts from canvas and non-canvas version. On the left side the keywords detected in Q1 are displayed, while on the right side they correspond to Q2.

Upon analyzing the detected keywords, it was found that Q1 received a higher number of keywords compared to Q2. In total, Q1 collected 242 keywords, while Q2 received 153 keywords (89 keywords fewer than Q1). This observation aligns with the nature of Q1, which delves into more personal experiences, while Q2 focuses on practical aspects. Visual graphs displaying these results are presented in Figure 19 and 20.

When comparing the canvas and non-canvas versions, the canvas version once again received a higher number of keywords. The canvas version collected a total of 206 keywords, while the non-canvas version received 189 keywords, a difference of 17 keywords.

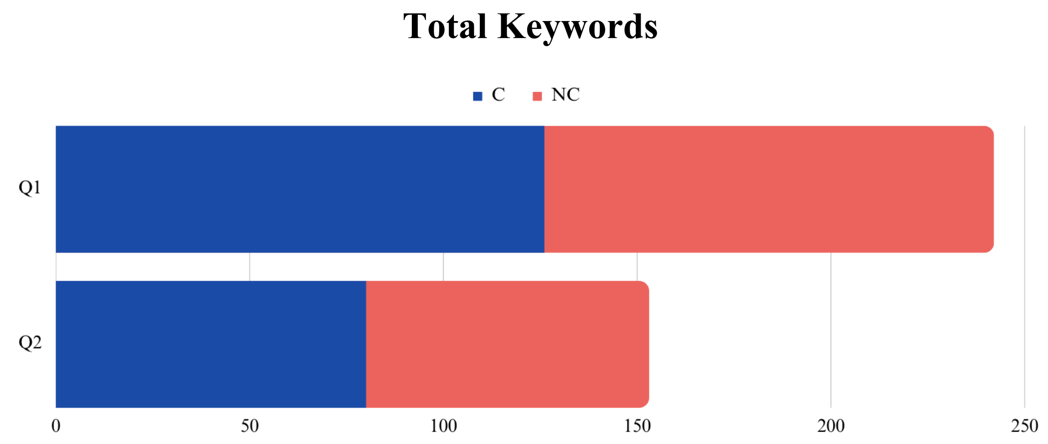


Figure 20. Number of keywords in percentages for contrasting canvas and non-canvas version.

Specifically, in Q1, the canvas version received 126 keywords compared to 116 keywords in the non-canvas version (a difference of 10 keywords). In Q2, the canvas version received 80 keywords, while the non-canvas version obtained 73 keywords, a difference of 7 keywords.

On average, Q1 responses contained 2.3 keywords, whereas Q2 responses had an average of 1.4 keywords. In the canvas version, the average number of keywords was 2.3 in Q1 and 1.5 in Q2, while in the non-canvas version, it was 2.2 in Q1 and 1.4 in Q2.

6.1.5 Number of Suggestions

In addition to examining character count and keyword analysis, the number of suggestions provided by participants in response to Q2 offered another variable to assess the richness of information obtained. It is important to note that Q2 specifically asked participants to mention ways in which their institution could better support students, making the number of suggestions directly indicative of the information's richness. These results can be observed in Fig. 21.

A total of 161 suggestions were received, with 86 provided in the canvas version and 75 in the non-canvas version (a difference of 11). The average number of suggestions across all observed responses was 1.5. Notably, the canvas version exhibited a slightly higher average of 1.6 suggestions, while the non-canvas version had an average of 1.4 suggestions.

The maximum number of suggestions provided by a participant in the canvas version was 5, while the non-canvas version had a maximum of 4 suggestions. It is noteworthy that both versions also included responses without any suggestions. Specifically, in the canvas version, 7 responses lacked suggestions, while in the non-canvas version, there were 9 responses without any suggestions.

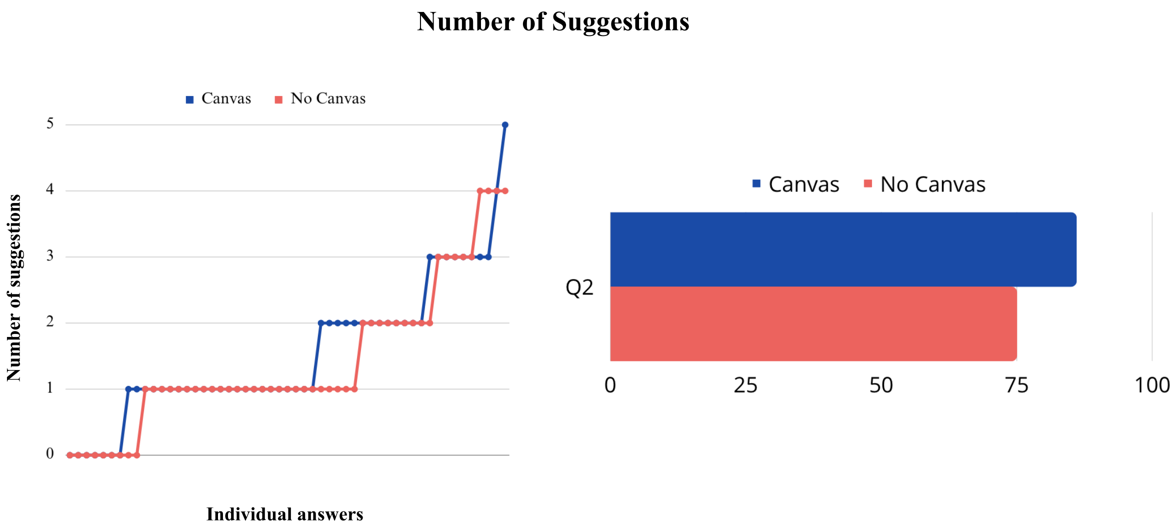


Figure 21. Number of suggestions analysis. On the left, they are displayed separated into canvas or non-canvas version, while on the right side, suggestions are displayed in percentages.

6.2 Experiment 2: Qualitative Results

6.2.1 Perceptions

Overall, the participants reported a positive experience with the survey, with 19 out of 37 participants mentioning that they liked, loved or enjoyed answering it. Moreover, 14 participants acknowledged the importance and relevance of the survey topic, while 5 participants found the survey and/or the topic interesting. Furthermore, 11 participants noted that the survey prompted them to reflect on the topic or think more deeply about their experiences. No negative comments about the overall experience were received.

In terms of how easy or difficult it was to answer the survey, 11 participants stated they found it easy or straightforward. However, opinions were divided when it came to how specific the questions were. Seven participants suggested that the survey could have included more specific questions, while 13 participants found the questions to be adequate, concrete, direct, and/or clear.

Regarding the length of the survey, 18 participants considered it to be relatively short. Again, opinions were divided on this aspect. Seven participants expressed a negative connotation associated with brevity, with comments such as:

> I did not go into much detail as I did not expect it to be so short

> It could have included more specific closed questions

> It was too concrete. Normally surveys have more questions that allow to answer different aspects of the topic

However, associating brevity with a negative aspect does not imply that they did not like the survey, as several of these participants also mentioned that they enjoyed answering it. On the other hand, out of those 18 participants, 11 perceived the brevity of the survey positively. They appreciated the concise nature of the questions, commenting:

> It was short but comprehensive

> I appreciate short and concrete surveys

> Excellent with sufficient, concrete, and to-the-point questions

> I loved that it was short and simple, with clear questions

Similarly, two participants emphasized that the open-ended format was more appropriate due to the personal nature of the topic, making it challenging to categorize it into standard topics for closed questions. Additionally, two participants mentioned that although they found the survey to be short (in a negative way), if it had been longer, they may not have provided as much detail. This contrasts with the participant mentioned in the previous paragraph who mentioned that the brevity of the survey led them to provide less elaboration.

6.2.2 Interpretations

Participants generally expressed hesitation when associating meaning with the generative data representations, often using phrases such as 'I think that...', 'to me it means...', 'I feel that...', and 'I thought it was...'. While none of the participants explicitly stated that the representations corresponded to the data representations of their written responses, the majority of participants (34 out of 37) did link meaning to their responses, primarily in terms of interaction.

These interpretations can be categorized into three levels. At the first level, 11 participants associated the representations with the length of their responses. They made comments such as:

> The more you write, the bigger it gets

> I noticed that the color changed more as I wrote more

> I thought it was indicating that I wrote enough

On that note, some participants associated the representations with the idea of meeting a character requirement, commenting on observations such as:

> I assumed it was related to the length of the answer and changed to indicate I reached a character limit

> Maybe it changed because I wrote enough, like when passwords turn green on a website

> I think it means that my answer is valid or that I reached a sufficient number of words

In this same level, another category of participants associated the representations with their progress or the sense of accomplishing a goal. They made comments like:

> I thought it was a positive sign that I wrote so many words

> To me, it meant that I was making progress and completing my response

> I think it likes it when you write a lot of words

At a second level, 12 participants associated the representations with the content of their answers, but on a general level. They perceived a connection between specific keywords and the changes in the representations. Some comments from this group include:

> I think that when I write a specific word, it changes, but the more I write, the more it grows

> Sometimes I deleted a word, and it changed; I think it is related to the words, but I don't know the exact connection

> The circle gives me the impression that it relates to the number of characters, while the background and other elements relate to keywords in my answer

> I think that when I mention a keyword, the program identifies it and changes the color or adds shapes

Finally, in a third level, 11 participants associated meaning with the semantic content of their responses beyond just keywords. They believed that the representations reflected the meaning conveyed in their responses. Comments from this group include:

> I assume it is related to the kind of emotions I wrote

> I think the colors are related to the positive and negative aspects I wrote, although they are very similar, and I'm not entirely sure what it means

> I think it has to do with what I wrote in my answer; with certain topics, it changed more, although the dots always appeared

> I think it represents what I wrote because it changed to purple precisely when I discussed the negative aspects

> I'm not sure, but I think it has to do with the theme of how university affects mental health, although I couldn't get the exact logic behind the changes

> I think it is a recognition of the text because it changed suddenly from one second to another

> I imagined that it is related to the content of what I was writing; if I discussed positive things, it took on certain colors and shapes, and if I mentioned negative things, it changed to others... I'm not sure if it's possible though

> Maybe it represents moods; at least, the gradient gave me that impression

6.2.3 Interaction Patterns

During the interviews, several interesting points and patterns emerged regarding participants' interactions with the generative data representations. Only 6 participants stated that the representations did not influence their answers or the way they responded, while the remaining 31 participants identified, to varying degrees, some form of influence.

Among the patterns that can be extracted from the latter group, positive influences were observed, where participants felt more motivated to complete their answers, provide more descriptive content, or elaborate further. Comments supporting these observations included statements such as:

> Maybe it motivated me to write more to understand what changes it could create

> I felt that it changed the background based on the number of characters, so I wrote a little more to see the background color change again, although I don't think it affected the content

> It motivated me to keep writing. The design was nice, so I thought, 'Now, instead of including 2 examples, I'll include 3 to see how the circle gorws'

> I think I went into more detail to see more changes. Although I also felt that due to the topic itself, I naturally wanted to explain it in depth

> I think it influenced me to use more specific terms or be more descriptive to see if it would trigger more changes

Participants also expressed heightened eagerness to take part and curiosity. They engaged in certain behaviors such as deleting words or experimenting before finalizing their responses. Comments illustrating these patterns included statements such as:

> Sometimes I deleted a word to see if it would trigger a change, and then I rewrote it

> As you write, you want it to be fuller. It is visually pleasing

> As I expressed myself more, it improved, so sometimes I wrote more out of curiosity

> I was intrigued by how the background changed as I wrote, particularly in the last question, as the previous one had already surprised me, and I was more expectant in the last one

> It didn't influence my answer, but I erased and rewrote just to see how it changed... experimenting... and then I wrote my original and official answer

> More than influencing my answer, it motivated me to provide a more comprehensive response or think of something more detailed

Moreover, participants associated the generative data representations with a sense of reward or incentive. Comments reflecting this perspective included statements such as:

> I felt like it was a little reward for providing more detail in my answers... maybe that was the objective?

> I wanted to see it fully filled

> I felt that as it was changing, I wanted to write more to complete certain parts as if it were a game

> It's fun to see the background change, both the dots and the colors added, it makes you want to write more

> It adds interactivity and increases the desire to answer

> The first time, it was entertaining to see how the figure was changing... after a while, I lost track, but when it changed to purple, it was like 'wow!... another call for attention', like if it was saying 'we're still listening to you'

> The second time, I already knew what would happen, so it wasn't as surprising, but I still made an effort to keep writing to reach the color change, to see if it would be the same or another visual transformation

However, some participants associated the generative data representations with negative aspects or boundaries of the system. They felt that a maximum had been reached or that, after a certain point, it was unnecessary to provide further details in their responses. Comments reflecting these perspectives included statements such as:

> When it turned red, I thought I had reached the maximum number of characters. It influenced me because I didn't know if I had to write more or be more concise

> At the beginning, I thought it was a warning that I had written too much and needed to stop

> I don't think the color influenced my response, but maybe I stopped writing because I felt that I had already achieved the goal of triggering a change

> I think it didn't influence the content, but it did influence the length. Although not significantly, I felt that I had written enough as I saw the representation change

Furthermore, certain usability issues were discovered during the interviews, which were relevant to interpreting the study results. Participants mentioned distractions or difficulties in readability. Regarding distractions, some participants commented:

> I got startled when the color suddenly changed in question 2

> The movement of the generative data representation was distracting

In terms of readability, participants also mentioned:

> There were some colors that bothered me because I couldn't see what I was writing

> What I was typing wasn't readable anymore

> My answer was partially hidden... since I answered on my cell phone, I'm unsure how it would appear on a computer screen. Also, after a while, the 'mandala' went off-screen

6.3 Summary of Findings

The findings from both Experiment 1 and Experiment 2 suggest that creative computing methods, specifically generative data representations, can indeed enhance participant engagement in qualitative research within the context of citizen science.

In Experiment 1, the quantitative analysis revealed that the use of generative data representations in the survey led to an increase in the average character count, keyword count, and the number of suggestions provided by participants. This suggests that the visual representations encouraged participants to provide more detailed and comprehensive responses, thereby enriching the collected data. The canvas version of the survey, which included the generative data representations, consistently yielded slightly richer and longer responses compared to the non-canvas version.

However, it is also important to recognize that while the canvas version showed promising results, these improvements were slight, indicating that further investigation and optimization are needed to fully understand and utilize the potential of generative data representations in online surveys.

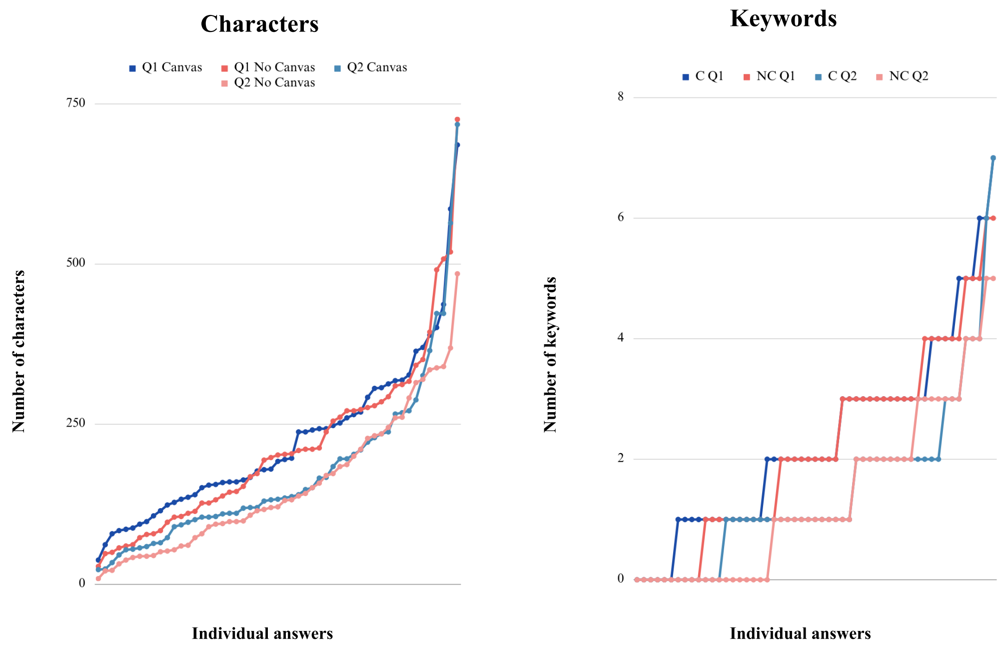


Figure 22. Summary of results. On the right side by number of characters, and by keywords detected on the left side. Each dot is an individual answer, sorted on the X axis in an ascendant way.

In Experiment 2, the qualitative feedback from participants further reinforced these findings. Participants generally reported a positive experience with the survey, with many expressing that the generative data representations made the survey more engaging and encouraged them to provide more detailed responses. Some participants even associated the changing representations with a sense of incentive or motivation, indicating an element of reward that could potentially enhance participant engagement in citizen science projects.

However, it is important to note that while the generative data representations generally had a positive impact on participant engagement, some participants reported usability issues such as distractions and difficulties in readability. This suggests that careful consideration must be given to the design and implementation of these methods to ensure they do not inadvertently hinder the user experience, or generate biases in the results.

In summary, the findings from both experiments provide evidence in support of the research question and hypothesis. They suggest that creative computing methods, specifically generative data representations, can be effectively utilized to augment tools for qualitative research in citizen science. These findings set the stage for a more in-depth discussion of these results and their implications, which will be explored in the subsequent discussion chapter.

6.4 Discussion

The results of both experiments present interesting findings regarding the impact of generative data representations on participants' engagement and response behavior. The findings from Experiment 2 offer insights that help elucidate some of the patterns—regular and irregular—identified in Experiment 1. For instance, a possible explanation for why the response curve exhibits clear behavior for responses below the mean character count, yet less predictability for those above the mean, can be derived from these insights. The following discussion elucidates these connections in greater detail.

6.4.1 Engagement and Anticipation

The increase in character count, keyword count, and suggestions observed in Experiment 1 results was consistent with participants' reports in Experiment 2, indicating an increased motivation to compose more detailed responses. Apparently, this motivation was due to the fact that participants found the visual changes to be a form of reward or incentive, which encouraged them to participate mindfully in the survey.

This might clarify why the difference between results from the canvas and non-canvas versions was smaller for Q1 compared to Q2. The canvas version of Q1 presented an average increase of 9 characters, while for Q2 was 20. The increases on the canvas version over Q1 and Q2 answers can be seen in Figure 23, highlighted in green.

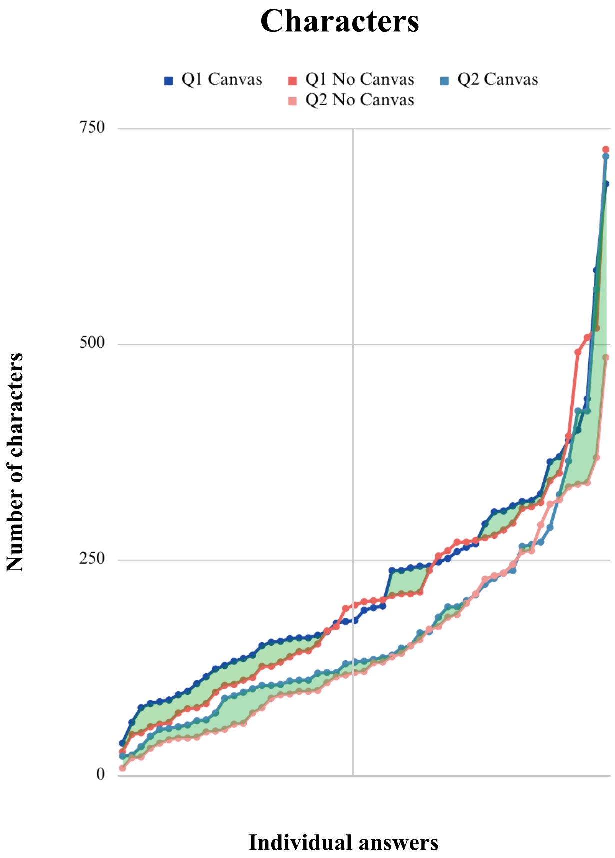


Figure 23. Character count for all answers. As it can be noticed, the trend is clear on the left half of the graph and becomes irregular on the right half, where answers have more characters.

Based on feedback from participants who interacted with the canvas version, generative data representations were initially a positive surprise during Q1, leading to heightened anticipation when facing Q2. Participants approached it more expectantly to find out changes originated by their answers. Conversely, those who encountered the version without canvas were presented with a standard survey. Literature suggests that participants' motivation tends to decrease throughout the surveys, leading to shorter and less comprehensive responses (Galesic & Bosnjak, 2009).

6.4.2 Rewards and Intrinsic Motivation

Some participants associated generative data representations with rewards or incentives, which links to the theories of motivation and gamification (Zichermann & Cunningham, 2011). For some, the shifting representations seemed to act as feedback for participants, indicating their progress or accomplishment. This feedback mechanism potentially ignited intrinsic motivation, fostering a sense of competence, auto-nomy, or enjoyment.

This trend of increased engagement through visual feedback is more apparent in the first half of the graph, that is, in responses below the average number of characters, as presented in Figure 23. Here, responses in the canvas version are slightly but consistently longer, contain more keywords, and offer more suggestions. In the interviews, some participants even commented that they increased the number of suggestions they gave to witness more changes in the representations. This observation aligns with the graph of the results of Experiment 1, and can be seen in more detail in Figure 24. In the non-canvas version, the average number of suggestions was 1.42, which rose to 1.59 in the version with canvas.

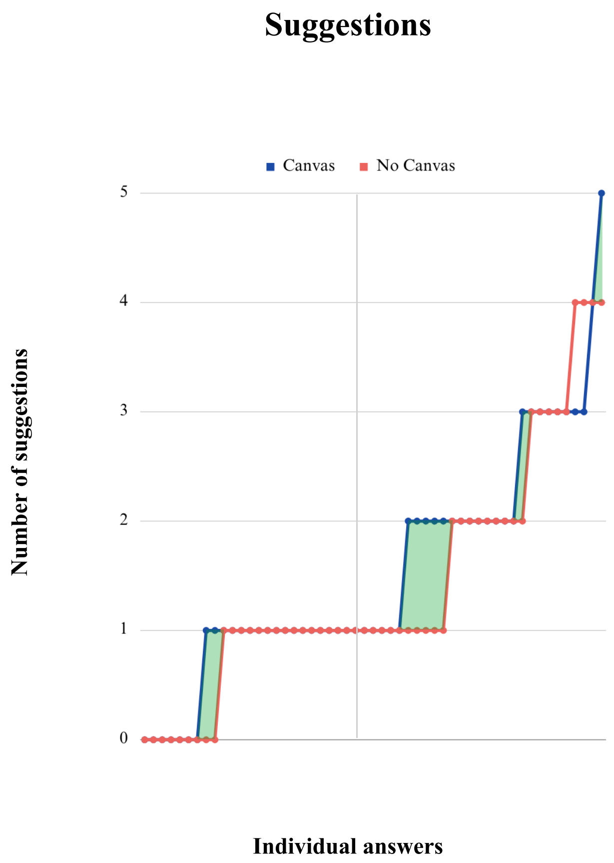


Figure 24. Number of suggestions by answer for Q2.

Building on these observations, curiosity emerged as another strong intrinsic motivator. A number of participants expressed curiosity about how their responses could alter the representations, thereby influencing not only their willingness to respond but also their approach and response content. This finding echoes literature that underlines the power of curiosity as a potent intrinsic motivator (Oudeyer et al., 2016).

6.4.3 Negative Influence

Despite the overall trend of enhanced engagement, the generative data representations did not positively influence all participants. Some participants remained unaffected or their experiences were even negatively affected. Possible explanations could include individual differences such as personal preferences or cognitive processing styles, as these variables might lead people to interpret the interactions in different ways.

For example, some participants who believed they had reached a maximum character limit or did not feel the need to provide additional details after a certain point, may have interpreted the visual cues as indicators of completion.

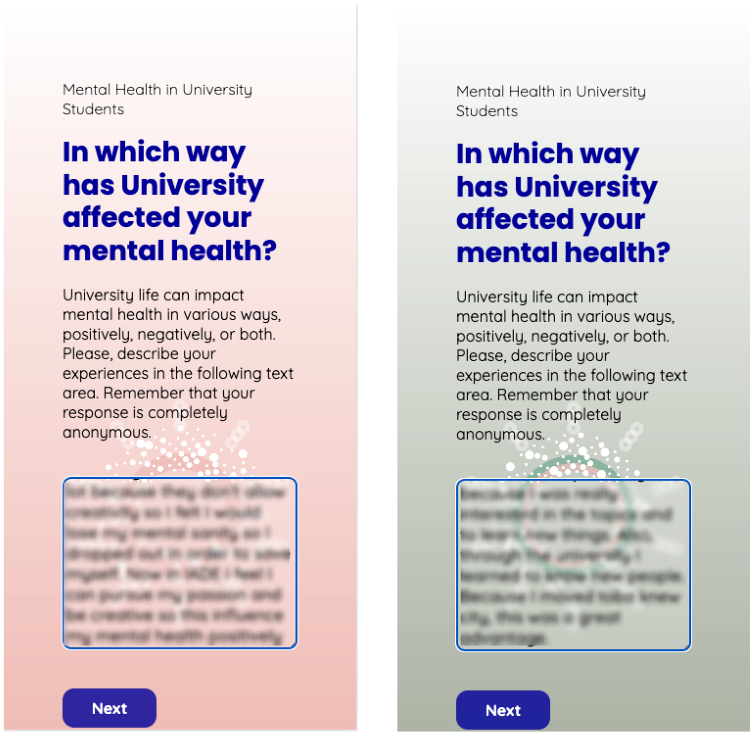


Figure 25. Interfaces of two participants on their actual answers. On the left side, the color was associated to a limit on the amount of characters permitted, while on the left side the participant felt green meant that it was complete enough to be submitted.

This assumption about reaching a maximum is particularly relevant for participants who interpreted the colors red and green as signals to stop. Their interfaces can be seen in Figure 25. In these cases, they had already composed an answer of a certain length when the color changes occurred, which could explain the second part of the graph. For responses with more characters than the average, the opposite effect is noticeable at certain points, with the non-canvas version of the survey yielding more detailed data than the canvas version. Participants who had already composed what they thought was an acceptable amount of characters might have been more inclined to interpret a color change as a signal of goal completion or reaching a limit. The possible negative effects of the system are highlighted in Figure 26.

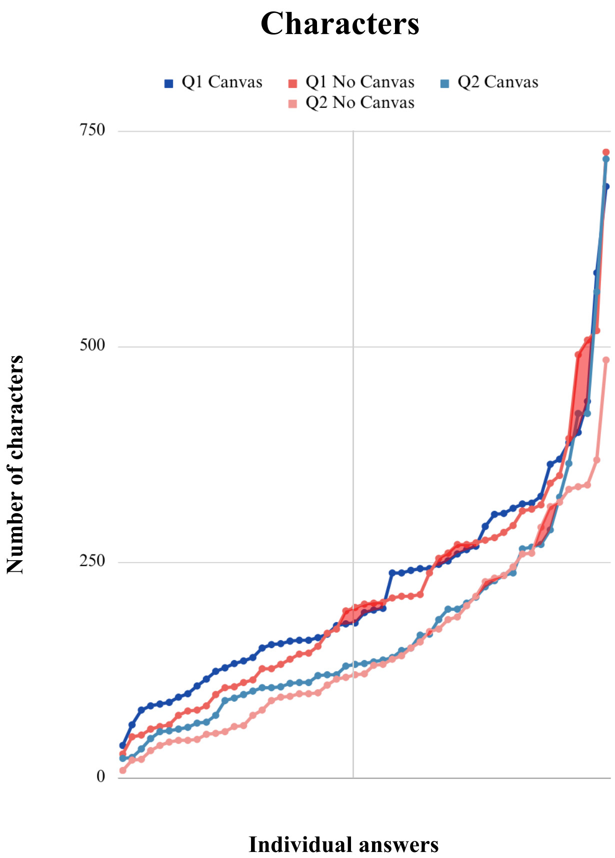


Figure 26. Character counts for all answer contrasted. In red it is highlighted the cases where the non-canvas version outperformed the canvas version.

These interpretations underscore the importance of clear communication about the purpose and functioning of the generative data representations to avoid misconceptions and foster comprehensive responses. More research should be done to understand how managing expectations and effectively communicating the purpose of the representations could affect engagement.

6.4.4 Usability Issues

A variety of usability issues were also identified, including concerns about potential distractions and readability difficulties. Some participants found the abrupt changes in color or movement to be startling or distracting, suggesting that the design of such data representations needs to carefully balance between promoting engagement through interactivity and avoiding unnecessary distractions. Furthermore, readability issues underscore the need to consider the variability of screen sizes and devices during the survey design process. Two interfaces illustrating these issues can be seen in Figure 27. Employing user-centered design principles and iterative testing could significantly aid in identifying and mitigating such usability concerns prior to survey deployment.

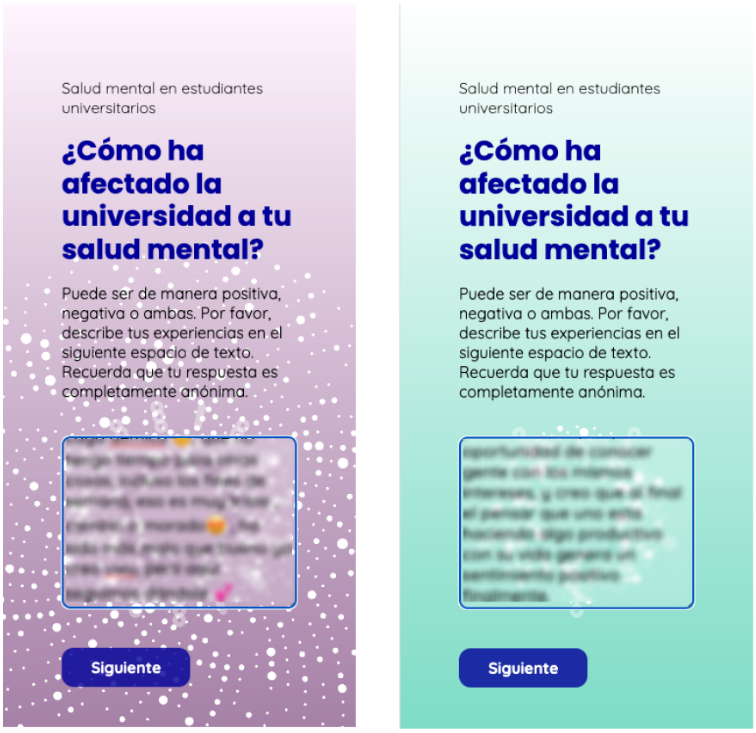


Figure 27. Two interfaces with reported usability issues. On the right side the continuous feedback loop had reached the limits of the screen so it was not visible anymore, while on the right side the participant complained that they could not read the answer due to the colors.

6.4.5 Summary of Discussion

The findings from both experiments reveal a complex interaction between generative data representations and participant engagement. Positive trends include increased response length and more detailed content, driven by perceived rewards and intrinsic motivation. However, the data also reveal potential obstacles to engagement, such as negative interpretations of color changes and usability issues. Understanding these influences is key in creating more effective and engaging surveys using generative data representations.

7 Conclusions

The study's primary research question, *Can creative computing be utilized to generate or augment tools for qualitative research in citizen science, prioritizing participant engagement?* was addressed through empirical experiments and thorough analysis of participant interactions and responses. The study demonstrated that creative computing methods, specifically generative data representations, could indeed bolster engagement and enrich data in the context of citizen science research.

The objectives of designing and developing creative computing tools specifically designed for qualitative data collection within citizen science projects were met by the novel implementation of generative data representations. As for other objectives, evaluating and assessing the influence of them in participants' answers quality and experience, they were met through the design and execution of the experiments.

The outcomes of Experiment 1 and Experiment 2 validated the hypothesis that employing generative data representations can increase participant engagement in qualitative research in the scope of citizen science projects. The dynamic and interactive nature of the generative data representations served as a motivator, prompting participants to give more detailed and thoughtful responses. A segment of participants even associated the continual changes in data representation with a sense of reward or incentive, illustrating the great potential of creative computing in research methods.

7.1 Contributions

The main contributions of this study are:

The introduction of a novel approach to an existing problem, shedding light on the value of meaningful interactions.

The validation and promotion of creative computing as an applied discipline, demonstrating its potential to enrich qualitative research and enhance participant engagement in citizen science.

The creation and implementation of 'generative data representations', a visual representation system for qualitative data that can have multiple applications.

The proposal and exploration of non-traditional methods for engagement and gamification, breaking away from conventional strategies.

Part of this research was presented at the Expat'23 conference (Lagos Rojas et al., 2023), contributing to the ongoing discourse in the respective field.

7.2 Implications

This section delineates the implications that emerge from the study's findings. These implications are of interest not only to the fields of creative computing and citizen science but also to broader contexts where participant engagement is essential.

7.2.1 Potential to Enhance Data Richness

The results from this study present several implications, particularly for citizen science and qualitative research. A significant finding is the potential of generative data representations to enhance participant engagement in surveys. By employing these creative computing methods, character count, keyword usage, and the number of suggestions received increased. Moreover, the qualitative insights revealed an increased motivation among participants to provide more detailed responses when presented with these generative data representations. Hence, by integrating these creative computing techniques into their methodologies, researchers and citizen scientists can enrich their data collection process, potentially yielding more comprehensive and insightful responses.

7.2.2 Interactivity trade-off

Participants' feedback highlighted concerns regarding potential distractions and readability issues, emphasizing the need for strategic design in implementing generative data representations. To strike an optimal balance between engagement and readability, the design of these type of visual cues must be thoughtful and consider potential cognitive biases. They should serve to enrich the participant's experience rather than hinder it. Furthermore, consideration of diversity in user devices, screen sizes, and preferences during the design phase can significantly enhance the user experience and accessibility of such techniques. Consequently, a user-centered approach to design that prioritizes user experience and accessibility is essential when implementing these techniques.

7.2.3 Other Means for Gamification

The study further provides insights into how generative data representations can expand the perception and implementation of gamification techniques within citizen science. While the approach does not employ traditional gamification elements, such as points and badges, it aims to augment participants' intrinsic motivation through elements like curiosity and surprise. Participants associated the dynamic visual changes in response to their inputs with rewards or incentives, demonstrating that such elements can operate as positive reinforcement within a research setting. This opens up the potential for utilizing these creative interactions as nontraditional gamification elements, which could effectively motivate and sustain participant engagement in online surveys and similar participatory platforms. In this way, this study broadens the perspective on what constitutes gamification.

7.2.4 Research Tools by Citizen Scientists

These findings have implications for researchers and citizen scientists involved in designing research tools. In citizen science, the design process should be rooted in a commitment to foster participant engagement and motivation, as these factors play a significant role in collecting rich and comprehensive data, but also in participant's experience and motivation with the project.

Importantly, the system discussed in this study presents an advantageous alternative to traditional gamification techniques, which often require constant monitoring and adjustments (Zichermann & Cunningham, 2011). In contrast, this system is designed to be self-sustaining and does not necessitate continuous tweaking or manual intervention, significantly simplifying its implementation and ongoing management. Moreover, compared to other projects that have developed full mini-game approaches (Harms et al., 2015), the present system offers a less complex and more easily deployable solution. As such, it offers a promising direction for developing user-centered research tools that can be more feasibly implemented in various citizen science contexts.

Clear communication about the purpose and interpretation of visual cues is also paramount to prevent potential misconceptions. Furthermore, embracing user-centered design principles, including iterative testing and feedback sessions, can facilitate the identification and resolution of potential usability issues.

7.2.5 Summary of Implications

Implications are for both the realms of creative computing and citizen science. The strategic integration of generative data representations can boost participant engagement, and enrich data collection, which can contribute to a more nuanced understanding of the research phenomena. Furthermore, adopting non-traditional gamification elements and user-centered design principles can help to foster a more inclusive, engaging, and effective research environment. Therefore, the findings obtained from this study open new paths for the design and implementation of qualitative research tools using creative computing approaches.

7.3 Limitations

While this study provides valuable insights into the influence of generative data representations on participant engagement and the potential of creative computing methods in citizen science, several limitations must be acknowledged.

7.3.1 Participant Sample

One notable limitation lies in the size and composition of the participant sample. All of the participants were university students, from various disciplines and institutions, a selection based largely on convenience. While efforts were made to ensure diversity within this group, it might not be possible to generalize to a broader and more varied population. Additionally, participant recruitment proved to be a challenge and benefited from on-campus strategies, which might not be feasible with groups that do not share a physical location.

7.3.2 Survey Design

The design of the survey itself also represents a limitation. A myriad of factors might influence participants' responses in traditional surveys, including question-wording, survey length, question order, and layout format. Additionally, the use of data visualizations introduces several trade-offs, as emphasized in the book *The Functional Art*(Cairo, 2016). These include balancing familiarity with originality, abstraction with figuration, and unidimensionality with multidimensionality, among others. The comprehensive list of trade-offs is depicted in Figure 28. These considerations are crucial in the encoding of data into visual outcomes, as they can significantly shape participants' perceptions.

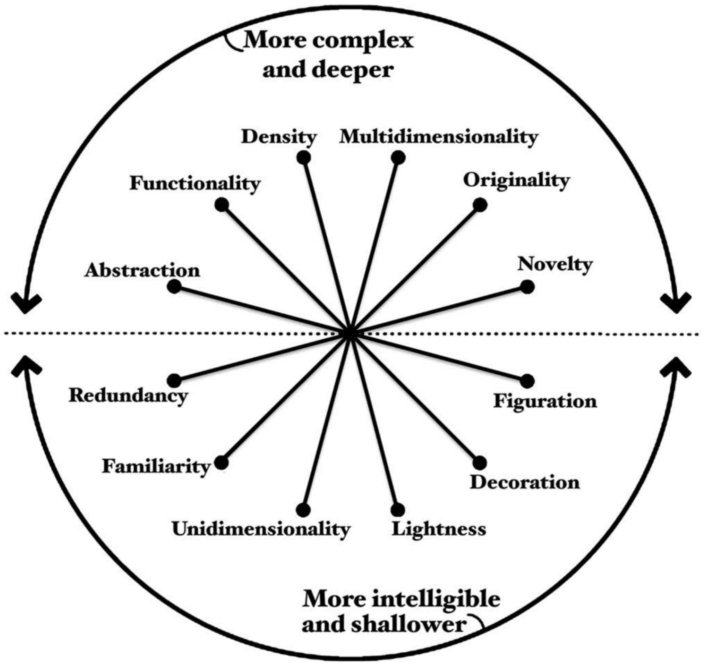


Figure 28. Visualization Wheel. It is a tool to reflect on the trade-offs of visualization design. Source: Cairo (2016).

7.3.3 Summary of Limitations

In conclusion, this study, while providing valuable insights into the impact of generative data representations on participant engagement, is bound by certain constraints. The limitations associated with participant demographics, survey design, and the inherent complexities of data visualization underscore the need for a nuanced understanding and cautious interpretation of the findings. These limitations also outline avenues for future research to address, potentially leading to a more comprehensive understanding of creative computing's role in citizen science and qualitative research.

7.4 Future Research

From these findings and limitations arise interesting paths for future research in the field of creative computing, citizen science, qualitative research, and information visualization, which are discussed in this subsection.

7.4.1 Addressing Limitations

Future research endeavors could address the participant diversity limitation highlighted in this study. Specifically, these studies could aim to include participants from broader demographic backgrounds, educational statuses, and cultural contexts. Applying the methodology employed here to different citizen science use cases across varied fields could enhance the findings' generalizability. The inclusion of diverse groups might uncover fresh insights about the ways these groups interact with generative data representations.

Additionally, subsequent research should investigate the impact of various design trade-offs on participant responses and their level of engagement. This exploration could include creating and testing alternative data representations, as well as examining how the length, order, format, and layout of survey questions influence participant interaction and response behavior.

7.4.2 Extending the Findings

Future studies should also extend the findings of this study. For instance, research could delve deeper into understanding the mechanisms that trigger visual changes, acting as a form of reward, and consequently enhancing participants' intrinsic motivation.

Considering the potential of generative data representations as incentives, future studies could also explore their effectiveness in facilitating recruitment processes. If the interactivity of generative data representations is presented to potential participants upfront, they might be motivated to participate out of curiosity.

Moreover, given that participants associated various meanings with the generative data representations, future research could further probe into these associations and their influence on response behaviors. Understanding whether certain colors or movements universally signify specific meanings, or if these associations are more personal, could offer valuable insights.

Lastly, the temporal aspect of presenting representations, specifically, the efficacy of real-time versus *a posteriori* displays, requires further exploration. While real-time representations increase interactivity, they can also potentially cause distractions or bisases. Conversely, presenting representations after participants submit their answers may reduce such distractions and potential biases, as participants are less likely to alter their responses based on real-time visual feedback. Studying the effectiveness of these different modes of presentation could offer novel insights into the optimal application of generative data representations.

7.4.3 Broadening the Scope

Future research could transcend the specific findings and limitations of this study, fostering a wider scope of inquiry. For example, subsequent investigations could assess the impacts of generative data representations across varied research contexts and explore their applicability in disparate fields such as marketing, policy-making, or healthcare. Moreover, it would be fruitful to determine whether the benefits of these methods endure across diverse domains.

In addition, there is potential to harness these methods for real-time data representation and feedback in dynamic contexts like participatory design workshops or brainstorming sessions. In these settings, immediate visual feedback could act as a catalyst for creative thinking and enhance collaborative problem-solving, thereby boosting the overall productivity of these sessions. Probing into these possibilities can broaden the understanding of the far-reaching potential of generative data representations.

Figure 29. Group of generative data representations from participants' answers. Some visual patterns are already noticeable. Although they're not intended for the purpose of extracting meaning, they present an opportunity.

Moreover, this study has not delved into the potential of data representations to facilitate qualitative analysis for researchers, a prospect that could be both intriguing and transformative. The results already show some emerging patterns in visual representations that indicate the possibility to group or extract some data dimensions, as can be observed in Fig. 29. Future research could focus on this opportunity potential, potentially impacting the way qualitative analysis of text data is performed.

7.4.4 Summary of Future Research

In summary, the proposed future research avenues aim to address the limitations identified in this study, extend its current findings, and broaden the overall scope of investigation. This includes the enhancement of participant diversity, further exploration of design trade-offs, a deeper understanding of the mechanisms of visualization-based rewards, and an examination of the application of generative data representations in diverse fields and contexts. Additionally, the potential for data representations to facilitate qualitative analysis emerges as an exciting new direction for research.

Ultimately, the continuation of research in this area can allow for exploring new frontiers in the integration of creative computing and qualitative research, contributing to the advancement and adoption of citizen science research, by providing more adequate yet creative tools for participant engagement.

7.5 Final Thoughts and Closing Remarks

This study presents a novel approach in integrating creative computing with qualitative research in citizen science, highlighting the considerable potential of such methods for fostering participant engagement and motivation. Despite certain limitations, the study paves the way for the further application of these techniques in citizen science and beyond, by enhancing qualitative research tools and facilitating the implementation of citizen science projects. The journey to refine these methods and explore new potentials at the intersection of creative computing and citizen science is ongoing, and this study marks a milestone along this path.

Reflecting on the process, expectations needed to be adjusted regarding the time required to conclude the study. Initially, a broader scope was envisioned, and the intention to to delve deeper remained even as the study advanced. The results are indeed interesting and motivating; however, the time constraint limited a more comprehensive exploration of how the system can serve as a recruitment incentive. It was particularly interesting to investigate how situated visualizations, made from the proposed system, could impact the recruitment process, tackling both issue with a unified solution.

However, such endeavors will be taken up in future work. During the research process, a collaboration was initiated with an important museum in Chile, which agreed to pilot the system in relation to experience surveys. This will enable visitors to share their experiences about specific exhibitions and create visual representations in the visual language of the exhibition. Thus, this study not only presents interesting applications for augmenting interactivity but also begins to explore the concept of data representations as a form of personalized incentive.

The exploration undertaken in this thesis underscores the rich possibilities that lie at the intersection of creative computing, citizen science, and qualitative research. The use of generative data representations has shown how such techniques can impact participant engagement, bringing a fresh perspective to the realm of citizen science and expand its potential reach.

References

Adona, N. (2017). Fred Hutch Visitor Center. Retrieved June 1, 2023, from https://segd.org/fred-hutch-visitor-center

Albert, A., Balázs, B., Butkevičienė, E., Mayer, K., & Perelló, J. (2021). Citizen Social Science: New and Established Approaches to Participation in Social Research. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 119–138). Springer International Publishing. https://doi.org/10.1007/978-3-030-58278-4_7

Almeida, C., Kalinowski, M., & Feijó, B. (2021). A Systematic Mapping of Negative Effects of Gamification in Education/Learning Systems. https://doi.org/10.1109/seaa53835.2021.00011

Amirrudin, A., Harrigan, N., & Naqvi, I. (2021). Scaled, citizen-led, and public qualitative research: A framework for citizen social science [Publisher: SAGE Publications Ltd]. *Current Sociology*. https://doi.org/10.1177/00113921211056057

Andrade, C. (2020). The Limitations of Online Surveys. *Indian Journal of Psychological Medicine*, *42*(6), 575–576. https://doi.org/10.1177/0253717620957496

Antoun, C., Couper, M. P., & Conrad, F. G. (2017). Effects of Mobile versus PC Web on Survey Response Quality. *Public Opinion Quarterly*, *81*(S1), 280–306. https://doi.org/10.1093/poq/nfw088

Ars Electronica. (2023). European union prize for citizen science. Retrieved June 1, 2023, from https://ars.electronica.art/citizenscience/en/

Balázs, B., Mooney, P., Nováková, E., Bastin, L., & Jokar Arsanjani, J. (2021). Data quality in citizen science. In *The science of citizen science* (pp. 139–157). Springer International Publishing.

Barrett, D., & Twycross, A. (2018). Data collection in qualitative research. *Evidence Based Nursing*, *21*(3), 63–64. https://doi.org/10.1136/eb-2018-102939

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next Steps for Citizen Science. *Science*, *343*(6178), 1436–1437. https://doi.org/10.1126/science.1251554

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving. *Converg. Int. J. Res. New Media Technol.*, *14*(1), 75–90.

Braun, V., Clarke, V., Boulton, E., Davey, L., & McEvoy, C. (2021). The online survey as a qualitative research tool. *International Journal of Social Research Methodology*, *24*(6), 641–654. https://doi.org/10.1080/13645579.2020.1805550

Braun, V., Clarke, V., & Gray, D. (2017). *Innovations in Qualitative Methods*. https://doi.org/10.1057/978-1-137-51018-1_13

Bruckermann, T., Sassenberg, K., Stillfried, M., Schumann, A., Brandt, M., & Harms, U. (2022). I'm fine with collecting data: Engagement profiles differ depending on scientific activities in an online community of a citizen science project. *Plos One*, *17*(10), e0275785. https://doi.org/10.1371/journal.pone.0275785

Cairo, A. (2016). *The truthful art*. New Riders Publishing.

Campos, R., Monteiro, J., & Carvalho, C. (2021). Engaged Citizen Social Science or the public participation in social science research. *Journal of Science Communication*, *20*(06), A06. https://doi.org/10.22323/2.20060206

Celino, I., Re Calegari, G., Scrocca, M., Zamorano, J., & Gonzalez Guardia, E. (2021). Participant motivation to engage in a citizen science campaign: The case of the TESS network. *Journal of Science Communication*, *20*(06), A03. https://doi.org/10.22323/2.20060203

Charvát, K., & Kepka, M. (2021). Crowdsourced data. In *Big data in bioeconomy* (pp. 63–67). Springer International Publishing.

Chou, Y.-K. (2015). *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. http://ci.nii.ac.jp/ncid/BB18977357

Christensen, A. I., Ekholm, O., Glümer, C., & Juel, K. (2014). Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. *European journal of public health*, *24*(2), 327–332. https://doi.org/10.1093/eurpub/ckt067

Civic Data Design Lab MIT. (n.d.). Digital matatus - civic data design lab. Retrieved June 1, 2023, from https://civicdatadesignlab.mit.edu/Digital-Matatus

Clarke, V., & Braun, V. (2013). *Successful qualitative research*. SAGE Publications.

Coenen, J., Biedermann, P., Claes, S., & Moere, A. V. (2021). The Stakeholder Perspective on Using Public Polling Displays for Civic Engagement. *C&T '21: Proceedings of the 10th International Conference on Communities & Technologies - Wicked Problems in the Age of Tech*, 61–74. https://doi.org/10.1145/3461564.3461585

Coenen, J., Houben, M., & Vande Moere, A. (2019). Citizen Dialogue Kit: Public Polling and Data Visualization Displays for Bottom-Up Citizen Participation. *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, 9–12. https://doi.org/10.1145/3301019.3325160

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

Cornelissen, E. (2021). Citizen Dialogue Kit design - Eline Cornelissen. Retrieved June 1, 2023, from https://www.elinecornelissen.be/citizen-dialogue-kit-design/

Couper, M. P. (2011). The Future of Modes of Data Collection. *Public Opinion Quarterly*, *75*(5), 889–908. https://doi.org/10.1093/poq/nfr046

Couper, M. P., & Peterson, G. J. (2017). Why Do Web Surveys Take Longer on Smartphones? *Social Science Computer Review*, *35*(3), 357–377. https://doi.org/10.1177/0894439316629932

Crain, R., Cooper, C., & Dickinson, J. L. (2014). Citizen Science: A Tool for Integrating Studies of Human and Natural Systems. *Annual Review of Environment and Resources*, *39*(1), 641–665. https://doi.org/10.1146/annurev-environ-030713-154609

Creswell, J. W. (2017). *Qualitative inquiry and research design* (4th ed.). SAGE Publications.

Daikeler, J., Bošnjak, M., & Manfreda, K. L. (2020). Web Versus Other Survey Modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of survey statistics and methodology*, *8*(3), 513–539. https://doi.org/10.1093/jssam/smz008

Davies, L., LeClair, K. L., Bagley, P., Blunt, H., Hinton, L., Ryan, S., & Ziebland, S. (2020). Face-to-Face Compared With Online Collected Accounts of Health and Illness Experiences: A Scoping Review. *Qualitative Health Research*, *30*(13), 2092–2102. https://doi.org/10.1177/1049732320935835

De Oliveira Salvador, P. T. C., Alves, K. Y. A., Rodrigues, C. C. F. M., & Oliveira, L. V. E. (2020). Online data collection strategies used in qualitative research of the health field: a scoping review. *Revista gaúcha de enfermagem*, *41*. https://doi.org/10.1590/1983-1447.2020.20190297

de Sherbinin, A., Bowser, A., Chuang, T.-R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A.,

& Sivakumar, K. (2021). The critical importance of citizen science data. *Front. Clim.*, *3*.

Diebold, F. X. (2012). On the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline, 6.

Digital Matatus. (n.d.). Digital matatus. Retrieved June 1, 2023, from http://digitalmatatus.com/

Domestic Data Streamers. (2021). You had to be a feminist. Retrieved December 21, 2022, from https://domesticstreamers.com/projects/you-had-to-be-a-feminist/

Druschke, C. G., & Seltzer, C. E. (2012). Failures of Engagement: Lessons Learned from a Citizen Science Pilot Study. *Applied Environmental Education & Communication*, *11*(3-4), 178–188. https://doi.org/10.1080/1533015X.2012.777224

Dubey, U. K. B., & Kothari, D. P. (2022). *Research Methodology: Techniques and Trends* (1st ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9781315167138

Edwards, P. G., Roberts, I., Clarke, M., DiGuiseppi, C., Wentz, R., Kwan, I., Cooper, R., Felix, L., & Pratap, S. (2009). Methods to increase response to postal and electronic questionnaires. *Cochrane Database of Systematic Reviews*, *2010*(1). https://doi.org/10.1002/14651858.mr000008.pub4

English, P., Richardson, M., & Garzón-Galvis, C. (2018). From Crowdsourcing to Extreme Citizen Science: Participatory Research for Environmental Health. *Annual Review of Public Health*, *39*(1), 335–350. https://doi.org/10.1146/annurev-publhealth-040617-013702

European Citizen Science Association (ECSA). (n.d.). European citizen science association. Retrieved June 1, 2023, from https://www.ecsa.ngo/

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, *26*(2), 132–139. https://doi.org/https://doi.org/10.1016/j.chb.2009.10.015

Galesic, M., & Bosnjak, M. (2009). Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Giannini, T., & Bowen, J. P. (2019). *Museums and Digital Culture*. Springer.

Global Citizen Science Partnership. (2022). Citizen science global partnership. Retrieved June 1, 2023, from http://citizenscienceglobal.org/

Gold, M. (2022). ECSA 10 Principles of Citizen Science [Publisher: Open Science Framework]. https://doi.org/10.17605/OSF.IO/XPR2N

Goodman, E., Stolterman, E., & Wakkary, R. (2011). Understanding interaction design practices. https://doi.org/10.1145/1978942.1979100

Grudin, J. (2005). Three Faces of Human-Computer Interaction. *IEEE Annals of the History of Computing*, *27*(4), 46–62. https://doi.org/10.1109/mahc.2005.67

Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. G. V. (2018). Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study. *Journal of Medical Internet Research*, *20*(6), e231. https://doi.org/10.2196/jmir.9702

Gura, T. (2013). Citizen science: Amateur experts. *Nature*, *496*, 259–261. https://doi.org/https://doi.org/10.1038/nj7444-259a

H. Davenport, T. (2014). How strategists use "big data" to support internal business decisions, discovery and production. *Strategy & Leadership*, *42*(4), 45–50. https://doi.org/10.1108/SL-05-2014-0034

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge* (pp. 105–122). Springer Netherlands.

Haklay, M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., & Vohland, K. (2021). What Is Citizen Science? The Challenges of Definition. https://doi.org/10.1007/978-3-030-58278-4_2

Hanson, R. H., & Marks, E. S. (1958). Influence of the interviewer on the accuracy of survey results. *J. Am. Stat. Assoc.*, *53*(283), 635–655.

Harms, J., Biegler, S., Wimmer, C., Kappel, K., & Grechenig, T. (2015). *Gamification of Online Surveys: Design Process, Case Study, and Evaluation*. Springer Science+Business Media. https://doi.org/10.1007/978-3-319-22701-6_16

Harrison, S., Henderson, J., Alderdice, F., & Quigley, M. A. (2019). Methods to increase response rates to a population-based maternity survey: a comparison of two pilot studies. *BMC Medical Research Methodology*, *19*(1). https://doi.org/10.1186/s12874-019-0702-3

Hart, A. G., Adcock, D., Barr, M., Church, S., Clegg, T., Copland, S., De Meyer, K., Dunkley, R., Pateman, R. M., Underhill, R., Wyles, K., & Pocock, M. J. O. (2022).

Understanding Engagement, Marketing, and Motivation to Benefit Recruitment and Retention in Citizen Science. *Citizen Science: Theory and Practice*, *7*(1), 5. https://doi.org/10.5334/cstp.436

Heerwegh, D., & Loosveldt, G. (2008). Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly*, *72*(5), 836–846. https://doi.org/10.1093/poq/nfn045

Hil, D., & Lachenmeier, N. (2022). *Visualizing Complexity*. Birkhäuser.

Interaction Design Foundation. (n.d.). Interaction Design - brief intro. Retrieved June 1, 2023, from https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/interaction-design-brief-intro

Kalleitner, F., Mühlböck, M., & Kittel, B. (2020). What's the Benefit of a Video? The Effect of Nonmaterial Incentives on Response Rate and Bias in Web Surveys. *Social Science Computer Review*, *40*(3), 700–716. https://doi.org/10.1177/0894439320918318

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 205395171663113. https://doi.org/10.1177/2053951716631130

Knowles, S., & Brown, P. H. (2021). Cash is not king in incentivizing online surveys. *Applied Economics Letters*, *28*(2), 105–108. https://doi.org/10.1080/13504851.2020.1734524

Krath, J., Schürmann, L., & Von Korflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, *125*, 106963. https://doi.org/10.1016/j.chb.2021.106963

Kuwamura, N., Fuyuno, M., & Yoshimura, R. (2021). Application of Gamification to Online Survey Forms: Development of Digital Template System "Bingo Survey" and Evaluation. *2021 Nicograph International (NicoInt)*, 62–69. https://doi.org/10.1109/NICOINT52941.2021.00019

LaDonna, K. A., Taylor, T., & Lingard, L. (2017). Why Open-Ended Survey Questions Are Unlikely to Support Rigorous Qualitative Insights. *Academic Medicine*, *93*(3), 347–349. https://doi.org/10.1097/acm.0000000000002088

Lagos Rojas, C., Estima, J., & Dias, J. A. (2023). Generative data representations to improve free-text answers in online qualitative surveys. *2023 6th Experiment International Conference (exp.at'23).*

Land-Zandstra, A., Agnello, G., & Gültekin, Y. S. (2021). Participants in citizen science. In *The science of citizen science* (pp. 243–259). Springer International Publishing.

Lee, K. A., Lee, J. R., & Bell, P. (2020). A review of Citizen Science within the Earth Sciences: Potential benefits and obstacles. *Proceedings of the Geologists' Association*, *131*(6), 605–617. https://doi.org/10.1016/j.pgeola.2020.07.010

Lemmens, R., Antoniou, V., Hummer, P., & Potsiou, C. (2021). Citizen science in the digital world of apps. In *The science of citizen science* (pp. 461–474). Springer International Publishing.

Lomas, A. (n.d.). Andy Lomas. Retrieved May 25, 2023, from https://andylomas.com/

Longo, L. (2020). Empowering qualitative research methods in education with artificial intelligence. In A. P. Costa, L. P. Reis, & A. Moreira (Eds.), *Computer supported qualitative research* (pp. 1–21). Springer International Publishing.

Lorenz, L. (2020). Addressing diversity in science communication through citizen social science. *Journal of Science Communication*, *19*(04), A04. https://doi.org/10.22323/2.19040204

Mavletova, A. (2013). Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review*, *31*(6), 725–743. https://doi.org/10.1177/0894439313485201

McFarland, D. A., & McFarland, H. R. (2015). Big Data and the danger of being precisely inaccurate [Publisher: SAGE Publications Ltd]. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715602495

Menon, V., & Muraleedharan, A. (2020). Internet-based surveys: relevance, methodological considerations and troubleshooting strategies. *General psychiatry*, *33*(5), e100264. https://doi.org/10.1136/gpsych-2020-100264

Moggridge, B. J. (2003). *Designing interactions.* https://doi.org/10.4324/9780203563298-15

of Positive Design, D. I. (2016). Emotion typology. Retrieved June 1, 2023, from https://emotiontypology.com/

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. J. (2016). *Intrinsic motivation, curiosity, and learning.* https://doi.org/10.1016/bs.pbr.2016.05.005

p5.js. (n.d.). p5.js. Retrieved June 2, 2023, from https://p5js.org/

Palau Robert. (2020). You had to be a feminist. Retrieved June 1, 2023, from https://palaurobert.gencat.cat/en/detall/noticia/Expo-feminisme

Preece, J. (2016). Citizen Science: New Research Challenges for Human–Computer Interaction. *International Journal of Human-computer Interaction*, *32*(8), 585–612. https://doi.org/10.1080/10447318.2016.1194153

Processing Foundation. (n.d.). Processing Foundation. Retrieved June 1, 2023, from https://processingfoundation.org/

Purdam, K. (2014). Citizen social science and citizen data? Methodological and ethical challenges for social research. *Current Sociology*, *62*(3), 374–392. https://doi.org/10.1177/0011392114527997

Rabhi, L., Falih, N., Afraites, A., & Bouikhalene, B. (2019). Big Data Approach and its applications in Various Fields: Review. *Procedia Computer Science*, *155*, 599–605. https://doi.org/10.1016/j.procs.2019.08.084

Research[x]Design - KU Leuven. (2022). Citizen dialog kit. Retrieved June 1, 2023, from https://citizendialogkit.org/

Rolstad, S., Adler, J. R., & Rydén, A. (2011). Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health*, *14*(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Sauermann, H., & Franzoni, C. (2015). Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences*, *112*(3), 679–684. https://doi.org/10.1073/pnas.1408907112

Senabre Hidalgo, E., Perelló, J., Becker, F., Bonhoure, I., Legris, M., & Cigarini, A. (2021). Participation and co-creation in citizen science. In *The science of citizen science* (pp. 199–218). Springer International Publishing.

Shiffman, D., Fry, S., & Marsh, Z. (2012). *The nature of code*. D. Shiffman.

Shiffman, D. (n.d.). The Nature of Code. Retrieved December 9, 2022, from https://natureofcode.com/

Silverman, D. (2021). *Qualitative research*. SAGE.

Stern, M. J., Bilgen, I., & Dillman, D. A. (2014). The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design [Publisher:

SAGE Publications Inc]. *Field Methods*, *26*(3), 284–301. https://doi.org/10.1177/1525822X13519561

Studio Matthews. (2022). Fred Hutchinson Cancer Research Center — Fred Hutch Visitor Center. Retrieved June 2, 2023, from https://studiomatthews.com/project/fred-hutch-visitor-center/

Tanweer, A., Gade, E. K., Krafft, P. M., & Dreier, S. K. (2021). Why the data revolution needs qualitative methods. *Harvard Data Science Review*.

Tauginienė, L., Butkevičienė, E., Vohland, K., Heinisch, B., Daskolia, M., Suškevičs, M., Portela, M., Balázs, B., & Prūse, B. (2020). Citizen science in the social sciences and humanities: The power of interdisciplinarity. *Palgrave Communications*, *6*(1), 89. https://doi.org/10.1057/s41599-020-0471-y

Thunberg, S., & Arnell, L. (2022). Pioneering the use of technologies in qualitative research – A research review of the use of digital interviews. *International Journal of Social Research Methodology*, *25*(6), 757–768. https://doi.org/10.1080/13645579.2021.1935565

Toepoel, V., Lugtig, P., Struminskaya, B., Elevelt, A., & Haan, M. (2020). Adapting surveys to the modern world: Comparing a research messenger design to a regular responsive design for online surveys. *Survey Practice*, *13*, 1–10. https://doi.org/10.29115/sp-2020-0010

Tourangeau, R., Sun, H., Yan, T., Maitland, A., Rivero, G., & Williams, D. F. (2018). Web Surveys by Smartphones and Tablets. *Social Science Computer Review*, *36*(5), 542–556. https://doi.org/10.1177/0894439317719438

Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: A survey. *Journal of Big Data*, *2*(1), 21. https://doi.org/10.1186/s40537-015-0030-3

Twis, M. K., Miller, V. J., Cronley, C., & Fields, N. (2020). Collecting qualitative data through mobile applications: A multiple case study report. *Journal of Technology in Human Services*, *38*(1), 38–53. https://doi.org/10.1080/15228835.2019.1599766

Vaughn, P., & Turner, C. (2016). Decoding via Coding: Analyzing Qualitative Text Data Through Thematic Coding and Survey Methodologies. *Journal of Library Administration*, *56*(1), 41–51. https://doi.org/10.1080/01930826.2015.1105035

West, S., & Pateman, R. (2016). Recruiting and Retaining Participants in Citizen Science: What Can Be Learned from the Volunteering Literature? *Citizen Science: Theory and Practice*, *1*(2), 15. https://doi.org/10.5334/cstp.8

Williams, S. (2020). *Data action: Using data for public good*. MIT Press.

Williams, S., White, A., Waiganjo, P., Orwa, D., & Klopp, J. (2015). The digital matatu project: Using cell phones to create an open source data for Nairobi's semi-formal bus system. *Journal of Transport Geography*, *49*, 39–51. https://doi.org/10.1016/j.jtrangeo.2015.10.005

Wu, M., Zhao, K., & Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. *Computers in human behavior reports*, *7*, 100206. https://doi.org/10.1016/j.chbr.2022.100206

Yang, H., & Zhang, L. (2016). Promoting Creative Computing: origin, scope, research and applications. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2016.02.001

Yin, S., & Kaynak, O. (2015). Big Data for Modern Industry: Challenges and Trends. *Proceedings of the IEEE*, *103*(2), 143–146. https://doi.org/10.1109/JPROC.2015.2388958

Zainuddin, Z. (2018). Students' learning performance and perceived motivation in gamified flipped-class instruction. *Computers education*, *126*, 75–88. https://doi.org/10.1016/j.compedu.2018.07.003

Zhang, L., Bowman, D. A., & Jones, C. (2019). Exploring Effects of Interactivity on Learning with Interactive Storytelling in Immersive Virtual Reality. https://doi.org/10.1109/vs-games.2019.8864531

Zichermann, G., & Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. https://ci.nii.ac.jp/ncid/BB08348570