




Article

How to Improve Impact Reporting for Sustainability

Vedran Lesic ¹, Richard E. Hodgett ^{1,*}, Alan Pearman ¹ and Amy Peace ²

¹ Centre for Decision Research, Leeds University Business School, Leeds LS2 9JT, UK; v.lesic@leeds.ac.uk (V.L.); a.d.pearman@lubs.leeds.ac.uk (A.P.)

² Britest Limited, The Innovation Centre, Sci-Tech Daresbury, Warrington WA4 4FS, UK; amy.peace@britest.co.uk

* Correspondence: r.e.hodgett@leeds.ac.uk

Received: 1 February 2019; Accepted: 13 March 2019; Published: 21 March 2019



Abstract: Measuring real-world impact is vital for demonstrating the success of a project and one of the most direct ways to justify taxpayers' contributions towards public funding. Impact reporting should identify and examine the potential positive and negative consequences of the continuing operations of a proposed project and suggest strategies to expand, further develop, mitigate, avoid or offset them. Designing a tool or methodology that will capture the impact of collaborative research and innovation projects related to sustainability requires input from technical experts but also from experts in the domains of survey design and communication. Without survey design insights and testing it can be very difficult to achieve unambiguous and accurate reporting of impacts. This paper proposes six key recommendations that should be considered for those monitoring projects when identifying metrics and designing a sustainability impact report. These recommendations stem from a series of in-depth interviews about sustainability and innovation impact reporting with research project co-ordinators in the process industries (e.g., cement, ceramics, chemicals, engineering, minerals and ores, non-ferrous metals, steel and water sectors). Our results show that factors such as ambiguous terminology, two-in-one questions, the stage of the project, over-hypothetical estimates, inadequate formats and alternatives and lack of guidelines can negatively influence the data collected in usual project monitoring activities and jeopardise the overall validity of the reporting. This work acts as a guideline for those monitoring to improve how they ask for impact data from projects, whether they are introducing new impact metrics or evaluating existing ones.

Keywords: impact reporting; sustainability; process industries; metrics; survey design; project monitoring

1. Introduction

Sustainability is high on the political agenda worldwide. More than 190 countries have recognised the need to accept greenhouse gas (GHG) emissions mitigation and adaptation policies [1]. Various stakeholder groups are putting pressure on organizations to become more sustainable [2–4] and governments are passing legislation requiring public disclosure of environmental data (e.g., the Netherlands, Japan, New Zealand: [5,6]). How successful the process of implementation is will depend to a large extent on how well it is understood by policy makers [7] and how well we can accurately demonstrate the impact of such policies [8].

Process industries, which are responsible for producing essential materials such as cement, glass, steel and chemicals, account for more than 50% of global industrial CO₂ emissions [9]. In an ideal industrial scenario, we would be able to measure a wide-range of sustainability metrics that would help us reshape existing policies and redirect efforts into areas that are falling behind. However, in reality, sustainability metrics are often ignored or under-reported in contrast to business-related metrics [10].

Process industries are vital for society, making the materials and products that we need (e.g., underpinning the construction and transport sectors). However, processing the raw materials

can be energy and resource intensive. Operating now as a contractual public-private partnership (PPP) under the EU Horizon2020 framework programme, the SPIRE initiative—Sustainable Process Industry through Resource and Energy Efficiency—was launched in 2012 as an alliance of eight sectors of the European processing industries (cement, ceramics, chemicals, engineering, minerals and ore, non-ferrous metals, steel and water). The SPIRE PPP is tasked with delivering projects that have potential to achieve:

- a reduction in fossil energy intensity of up to 30% from the levels then current by 2030;
- a reduction of up to 20% in non-renewable, primary raw material intensity compared to current levels by 2030;
- a reduction of greenhouse gas emissions by 20% below 1999 levels by 2020, with further reductions up to 40% by 2030 [1].

The sectors united under SPIRE include more than 450,000 individual enterprises, provide jobs for 6.8 million employees and generate annually more than € 1.600 billion in turnover. A. SPIRE refers to the European Association that manages and implements the SPIRE PPP and, with particular reference to this paper, also collates and publicises the outcomes of research and development projects funded by the EU under the SPIRE initiative. Each year, project coordinators are asked to complete an annual SPIRE PPP questionnaire about the impact of their project. SPIRE is contractually obliged to report and deliver on Key Performance Indicators (KPIs) by the European Commission. KPIs are a quantifiable measure used to evaluate the success of a programme in meeting previously specified criteria. These questionnaires are a means of getting the data to report on the KPIs. A. SPIRE and the European Commission (EC) use this data to inform future decisions related to similar initiatives and publish a report. Usually such reports are used for dissemination and communication of the results, training purposes, demonstrating the added value of the network to the EC and by all the partners for both internal and external evaluation and justifying the taxpayers' contributions [11]. The aim of this paper is to present six most common issues that should be considered for those monitoring projects when identifying metrics and designing an impact report that includes sustainability measures. In the following sections, we present an overview of literature related to impact reporting and introduce the current study.

1.1. Impact Reporting

Measuring real-world impact is important for demonstrating the success of a project and one of the most direct ways for justifying taxpayers' contributions, not only within SPIRE but more broadly. Impact reporting should identify and examine the potential positive and negative consequences of the continuing operations of a proposed project and suggest strategies to expand, further develop, mitigate, avoid or offset them [12]. The question of whether to report sustainability metrics is no longer an issue [13]. Manes-Rossi et al. [14] in their analysis on sustainable reporting showed there is already a high level of compliance by large European companies who are following the most recent European Union Guidelines about non-financial information. Therefore, the focus should be on reporting and how to demonstrate the impact accurately and consistently [15,16].

Reporting at company or organisation level, rather than project level, aims to focus more on informing decision-making process of external stakeholders [17], legitimisation [18], reputational enhancement [19] and marketing [20]. The common practice among companies is to produce annual reports composed of various indicators of progress and success [21]. KPMG in their International Surveys showed the development of sustainability practices of multinational corporations, including their policies, targets and features of their environmental management systems, since 1993 [22]. Reporting impact on the project level might require different types of metrics and analysis in order to demonstrate the progress of specific projects while reporting at the company level might use other methodologies in order to aggregate the impact of each of the projects to evaluate the overall progress of the organisation.

Whether at company or project level, there is still very limited agreement on how to go about the process of impact reporting, both the approach adopted towards identifying impact and the specific

metrics employed [17,23]. In real-world impact reporting, there is still no clear set of standards about what metrics should be reported. For example, there is no universal standard for the overall content of such reports, presentation format, external validation and other matters. Recognising this deficiency, recently there have been several efforts within SPIRE to address this challenge, through projects such as:

1. STYLE (“Sustainability Toolkit for easy Lifecycle Evaluation”) is a European project that focused on pragmatic tools for industrial project teams to evaluate the broader sustainability implications of making an improvement to a product or process [24]
2. MEASURE (“Metrics for Sustainability Assessment in European Process Industries”) is a European project created to generate an objective picture of the current state-of-the-art in the use of Life Cycle Assessment methods across process industries in Europe and to formulate key challenges and research needs [25]
3. SAMT (“Sustainability assessment methods and tools to support decision-making in the process industries”) is a European project that focused on increasing integration of sustainability assessment methods in decision making by reviewing and making recommendations about potential methods, tools and indicators for evaluating sustainability in the process industry [26]
4. SPRING (“Setting the framework for the enhanced impact of SPIRE projects”) is a European project that aims to enhance the uptake of new and novel systems and technologies for improving resource and energy efficiency in the EU process industries by focusing on the needs and barriers of industry-based decision-makers. It is the project from which this paper originates [27].

Other attempts have also been carried out via the Global Reporting Initiative, together with the United Nation Environment Program, to support organisations to better explain and articulate their impact in the domain of sustainability [17,28].

In parallel with the variety of views on what impacts to report and how, there is a correspondingly wide range of tools developed for reporting impact. These tools measure different aspects of sustainability and range from qualitative screening tools to online tools for assessing the social impacts of supply chains and through to quantitative assessment tools for carbon or water footprints [10,14].

1.2. Types of Metrics

Sustainability goes beyond just environmental factors and includes other factors such as economic and social. Projects that are developed to improve sustainability focus on improving one or more of these factors. Impact assessments of these projects are necessary to understand the potential scale of the improvements and the trade-offs required. SPIRE has KPIs defined by the European Commission to serve as metrics to track the progress of those projects. The objective of the KPIs is to highlight and serve as an example for evaluation guidelines of the project that will contribute to performing an impact assessment. The KPIs within SPIRE are divided into two principal groups to facilitate an impact evaluation: Operational KPIs and Sustainability, Innovation and Competitiveness KPIs.

The main focus of the operational KPIs is to evaluate SPIRE as an organizational and funding instrument and its suitability to facilitate the achievement of the SPIRE objectives. For example, operational KPIs include the percentage of proposals reaching the negotiation stage and measurement of the time from submission to project kick-off. The main focus of the Sustainability, Innovation and Competitiveness KPIs, on the other hand, is to evaluate SPIRE in terms of its global contribution to the macro eco-socio-economic picture connected to sustainability, industrial competitiveness and innovation. Examples include the number of patents registered and the number of project reporting evidence of (green) job creation and/or maintaining jobs. This set of metrics aims to provide a practical starting point for businesses across the EU and the world to improve the efficiency of their production processes and products, enabling them to contribute to sustainable development goals.

The European Environmental Agency (EEA) has categorised metrics into a three-level typology: (i) descriptive, (ii) efficiency and (iii) performance [29]. Descriptive metrics aim to explain the current situation. They provide the information, usually the absolute measure, without necessarily the need

for its interpretation, which might be difficult for evaluation purposes. Efficiency metrics address the question of how sustainably and efficiently resources are being used in the production of various goods. These metrics should contain information about the quantity of resource involved and about the quantity of productive output generated by that resource. Finally, Performance metrics compare current with desired conditions. They are usually defined relative to a desired objective or target that should be achieved throughout the project duration. Each of these metrics provides different information to funders [30].

Apart from this typology, there are other important perspectives to consider, such as the indicators proposed by the International Organisation for Standardisation (e.g., environmental, management and operational performance indicators; ISO, 2018). Taken as a whole, over 80% of all metrics reported are descriptive and the remainder is composed of performance (around 13%) and efficiency (less than 5%) metrics. However, such reporting can be time consuming and represents an additional burden on a project or may require specific data that might be difficult to report due to the nature of the project (e.g., energy efficiency indicators when there is more than one facility and more than one type of production taking place). In addition, it does not help that industry often tends to perceive sustainability as an after-thought when they are planning their projects, with technical and short-term economic aspects having priority [10,31]. Thus, sustainability reporting is often not built into project reporting and other requirements from the outset.

A further exacerbating factor is a commonly occurring tension between the need for funders to have early and definite indications of the success of their funding investment and the reality that many research projects take a considerable time to deliver their payoffs—a problem shared across many sectors and many funders, not just in the process industries. In the case of SPIRE projects this particularly refers to the Technology Readiness Levels (TRL). The TRL gives a precise idea of how close an innovation is to commercial operation (and thus pay-off). The SPIRE projects typically operate in the TRL 5–7 area, with some recent projects going into TRL 8. While such TRLs relate to work beyond the basic (laboratory) level, they are typically still some way from full evidence of large-scale practical application. All this suggests that there is a good case but also a real challenge for sustainability considerations to be embedded thoughtfully, explicitly and earlier in projects and their reporting cycle [32]. Among other things, this would require the identification and use of appropriate sustainability metrics and specific procedures for implementing the impact reporting process—a potentially complex and multi-disciplinary one. In this paper we are not focusing on the types of metrics themselves or the content of the sustainability metrics but rather on the way they are presented in impact reporting. By making impact reporting more consistent we expect that the process of measuring KPIs will be more effectively captured.

1.3. Background

This paper builds on the work carried out within the SPRING project and aims to identify the six most common issues that should be considered, especially by research funders, when developing sustainability metrics and designing an impact reporting procedure. These recommendations stem from a series of in-depth interviews about sustainability and innovation impact reporting with research project co-ordinators in the process industries but have wider applicability as well, across a range of industrial sectors. The paper strongly advocates that without survey design insights and testing, regardless of how carefully selected and well-defined the metrics used are, it will be tremendously difficult to achieve the unambiguous and accurate reporting of impacts that funders often require and user groups desire. The interviews were undertaken as part of the SPRING project with the aim of informing the design of future SPIRE PPP questionnaires.

The paper proceeds as follows: Section 2 focuses on the mental model framework and illustrates the research methodology. Section 3 presents the overall findings and discusses specific issues identified in the interviews. Section 4 provides the final recommendations for practitioners in this field.

2. Methods

2.1. Insights from Survey Design Literature

When trying to report the impacts of individual research projects, a potential problem is the use of terminology and concepts that are difficult for non-specialists to understand (e.g., civil servants, outside stakeholders or even project coordinators themselves if projects span a wide range of specialisms or if the required unit of reporting does not match the scale of the project itself). For example, engineers tend to use the expression “a 100-year flood” rather than talking about a flood that has a 1% chance of happening each year, without understanding that lay people might misinterpret this information so they believe that a flood will only happen on every 100th year exactly [33].

Thus, designing a procedure that will truly capture and convey the impact of a set of projects in all the ways that potential stakeholders may need requires an effort not only from the engineers who may be directly engaged in the project but also, for example, from experts in domains such as survey design and communication. An overall process that addresses such barriers is the mental models approach for developing more effective communications [34]. This approach begins by identifying what lay people should know about the given topic based on existing literature and recommendations from the experts. It then moves on to conduct in-depth interviews and survey techniques to elicit people’s mental models. After that the approach focuses on the differences between the expert and lay mental models. It aims to finish with an evaluation of proposed changes in communications to determine their effectiveness [35]. The importance of such approach has been demonstrated in improving economic surveys [36], user interface design [37] and e-learning materials [38]. For example, a study on peoples’ expectations of inflation showed that asking respondents about ‘inflation’ is perceived more difficult than asking respondents about ‘prices in general’ and can lead to lower number of responses [36]. In this paper we are primary focusing on think-aloud interviews that are an integral part of the wider mental model framework. Failing to report data accurately, misunderstanding the instructions or avoiding reporting difficult or unclear metrics might seriously jeopardise the existing efforts of projects like those funded through SPIRE and misinform future decisions made by stakeholders.

It is on this aspect of sustainability impact reporting that the current paper focuses. The survey design literature suggests that small variations in instructions, question wording, response format and question order may have a major effect on respondents’ interpretations of the questions as well as their responses (e.g., [39–42]). This literature treats questionnaires as a form of communication between researchers and respondents [41]. In order for this communication to be effective, respondents should be able to understand the questions as intended by the researchers and be given the opportunity to provide answers that reflect their beliefs and experiences, which in turn should be understandable to the researchers.

When designing a questionnaire, especially if the issues covered are unfamiliar, ambiguous or sensitive, the survey design literature recommends what are termed “think aloud” or “cognitive” interviews to examine whether respondents interpret questions as intended and how variations in their interpretations may affect their responses [43,44]. Think aloud interviews are open-ended semi-structured interviews in which people are asked to think out loud while answering survey questions. They are used to highlight *any* issue that arises during the interviews, not simply the frequency of those issues. Based on previous studies, it is estimated that 15 to 20 interviews are sufficient to identify the most common problems with a typical survey [35]. Think aloud interviews go substantially beyond the standard pilot survey in terms of the depth and richness of the insights provided. For example, Murphy et al. [45] successfully applied think aloud interviews to develop the new patient-reported outcome measure designed specifically for primary care called the Primary Care Outcomes Questionnaire. Similarly, Sooniste et al. [46] used this type of interview as a method to magnify the differences between liars and truth-tellers.

2.2. Think-Aloud Interviews

The questionnaire on which this study is based consisted of 31 questions across four different sections (see Table 1 for examples). Its aim was to measure environmental performance as well as economic and social impacts at project level, in coherence with the progress towards project-specific key performance indicators. A sample of 14 project coordinators from different organisations funded within the SPIRE network was interviewed. Participants were asked to think aloud while reading through the questionnaire material and answering the questions, providing detailed feedback about their experience. Because a project coordinator is the key person responsible for reporting the impact of their projects, we have focused on their experience with the questionnaire. In addition, however, we also conducted two interviews with the European Commission and SPIRE network management officials, in which participants discussed the importance of metrics for different stakeholders. This reflects a common challenge when reporting impact. There are many uses of and users of, the information, ranging from projects wishing to report their own outputs through to funders wishing to monitor progress; from projects seeking to convince outside stakeholders to take up their findings through to funders needing to convince their political paymasters that further tranches of funding for the relevant area of work are justified.

Table 1. Sample questions from Spire PPP Progress Monitoring—2017 Questionnaire.

#	Questions:	Issues Identified:	Specific Recommendations:
Q2	Which sectors of the process industry are involved in the project?	Open-ended questions are time-consuming	Providing options would decrease the response time
Q3	How many SMEs are involved in the project as partners?	No issue identified.	No recommendation needed.
Q4	Please provide one feature (a success story) to promote your project? Please indicate the feature shortly (2–3 lines) and suggest a person to contact for further information (if needed)	Not tailored to the stage of the project because early stage projects are less likely to have success stories	Adding “Not applicable/option”
Q5	Would you have further communication materials that can help us to better understand the impacts of the project (please, provide a link)?	Not applicable to every stage of the project.	Tailor the question to the stage of the project.
Q6	Will the project lead to launching any of the following into the market? (i) New product (goods, technology, service); (ii) new process; (iii) new method	Confusion among the options provided.	Offer options that are distinctive and additional examples for each of the choices.
Q9	Please, indicate markets concerned by the functions of the exploitable results described in Table 7. Indicate current market size (in EUR million) in the EU _____ and outside the EU _____.	Quite complex and hypothetical question.	Provide more guidelines on how to attempt to answer this question for specific project.
Q10	Please, describe the scenario of potential deployment for the exploitable results you assume by 2030. Please indicate expected market size and market share.	Quite complex and hypothetical question.	Ask for shorter time period and provide further instructions.
Q18	Do you foresee any policy barriers for the deployment of the project’s exploitable results in the involved SPIRE sectors and in other industrial sectors? Yes (please specify: _____). No.	Participants often reported that they don’t know the answer.	Adding “I don’t know option”
Q23	Are /will resource efficiency targets of the proposal be met? Yes Partially No If partially or no, please provide brief explanations why?	Potentially biased question and it does not apply for every project.	Ask first if the project has efficiency target set and if yes, potentially ask about the progress.
Q29	Does the project develop high-skilled jobs with new profiles? Yes No If Yes (please provide number of new high-skilled job profiles developed ____ and give one or two examples ____) If no, is it planned at a later stage of the project? Yes (When) No	Unclear concept definition.	Use appropriate language and rewrite the existing question.
Q30	Do you expect your project to have impacts in terms of job preservation or job creation? Yes (Please, give estimates _____) No	Two questions in one.	Separate the questions or only ask for one.

Specifically, the participants in the think-aloud interviews reviewed their answers to the 2017 SPIRE questionnaire. The four sections into which the questionnaire was divided were: (i) project identification and success stories, (ii) business impact, (iii) environmental impact and (iv) socio-economic impact. For the study, a specific interview protocol was devised that followed the structure of the 2017 SPIRE questionnaire, together with additional evaluation questions. The interviews lasted between 45 min to two hours. They were conducted over the telephone and recorded with verbal consent from the interviewee. Once the section of the interview that covered the survey questions was completed, participants were asked to provide more detailed feedback about the survey as a whole. In this way, we were able to capture detailed feedback and reflections about the existing questionnaire. The interviews were transcribed and the issues participants identified, both about the survey instructions and the survey questions themselves, were analysed. We have independently analysed each of the interviewees' transcripts by mapping the common themes and compiling a qualitative report. This was cross-checked by an independent judge for 20% of the transcript materials to evaluate the consistency and validity of the themes identified. Think-aloud interviews allowed us to identify all the issues that arose through the process of assessing the questionnaire. Based on those issues identified, we have developed a specific list of suggestions that should be considered in order to allow unbiased and efficient impact reporting. These recommendations go beyond the scope of this project and can support policy-makers when constructing any new tool to measure impact or for improving an existing one.

3. Results and Discussion

We initially discuss the general themes that emerged from the think aloud interviews and their implications for practice. Next, we discuss six specific issues that designers of sustainability impact and similar surveys should consider improving the quality and reliability of the responses they obtain from survey respondents.

3.1. Overall Findings

An effective survey design is the result of many decisions that need to fit together and support one another in a way that produces the most accurate and meaningful responses from respondents. The 2017 SPIRE questionnaire that was the focus of the interviews contained 31 questions and each of these was explored in detail with the participants. However, here we will only report the issues that were identified as the most common and that have broader implications for sustainability reporting beyond the immediate confines of the SPIRE exercise. These are the questions shown in Table 1.

The overall consensus of the project coordinators interviewed was that they would value a quantifiable tool to measure impact. However, they also reported multiple issues in relation to the 2017 SPIRE questionnaire. An initial, exploratory set of interview questions revealed that participants varied in the elapsed time they took to complete the questionnaire, from some respondents taking only 4 h up to others taking several days. Participants used their project proposal, previous deliverables, project reports, previous questionnaires, partners' support and internet resources to provide answers for the questionnaire. When asked about how to improve the questionnaire, participants requested:

- clearer explanations of existing terminology
- more question boxes and interval ranges
- more specific impact assessment for each exploitable outcome
- clearer guidelines in the research funding call while they were preparing the project.

Across participants there was no consensus as to what part of the questionnaire should be the key priority for impact reporting for process industries. Moreover, there were different views as between the European Commission and the indicators they highlight and the project coordinators' views and what they might see as important to report about the impact of their projects. For example, interviews with the European Commission underline the importance of environmental indicators while project coordinators were more focused on business indicators.

Current practice in impact reporting is to use an online questionnaire or tool to capture respondents' responses. This also applies to the 2017 SPIRE questionnaire. An important insight, strongly reinforced by our interviews, is that when developing a tailored survey design for such an online questionnaire, the specific features of on-line information gathering (both its advantages and the issues it potentially poses to respondents) need to be considered.

An online platform allows for more flexibility and creativity for both survey designers and respondents. For example, visual design elements can be of significant importance. How a survey is designed and administered should depend on the length and topic of the questionnaire [40]. Longer questionnaires on more complex topics such as reporting sustainability metrics or life cycle assessment indicators might require additional time for respondents. Particularly in this context when respondents are expected to provide a more quantitative response that requires additional effort prior to providing answers, the design and wording of the questions play an increasingly critical role.

The biggest areas of concern raised by both the European Commission and SPIRE was the accuracy of the data, confidence in the results and meaningfulness of the data interpretation. Conducting surveys that reflect the current situation in a project requires developing procedures that minimise different types of error, for example, error of measurement, coverage, sampling or nonresponse error [40]. Reducing survey error means selecting the survey mode or combination of modes that provides appropriate coverage of the specified population and designing an implementation system that encourages respondents to provide thoughtful and honest answers. Not having a full picture of how individual projects have performed on different metrics will jeopardise the ability to convey a true picture of the impact of the research funding initiative as a whole. In particular, if projects that do not share their data achieved only modest or low progress on sustainability metrics such as energy efficiency and resource use, there is danger of the work being under-appreciated.

Finally, the biggest threat for impact reporting is the measurement error that often occurs when a respondent's answers are inaccurate, a complete guess or not measuring what they are supposed to measure. This was particularly highlighted by project coordinators in the interviews. Constructing an on-line questionnaire involves making sure that items are displayed in the same way for every communication format such as different browsers, computers, tablets and other technologies. Keeping display consistent will ensure that measurement error is reduced to a minimum because the survey design literature shows that user-friendly questionnaires have higher response rates [40]. Later in this section, we address those errors of measurement by providing best practices to avoid such mistakes.

Pre-testing the questionnaire is one of the key guidelines for delivering a high quality and effective survey design [47]. The term pre-testing is usually mentioned with respect to conducting a questionnaire with individuals who have certain expertise on a topic or, indeed, any member of the population. This involves asking selected people to complete the questionnaire and report any problems they experience. But what is done can and in appropriate circumstances should, go well beyond simply completing the questionnaire and noting any issues that come to mind. There are real benefits to be had from deploying more sophisticated methods, such as think aloud interviews, behavioural coding, response latency, vignette analysis, formal respondent debriefings, experiments and statistical modelling [48]. Minimally, we would suggest obtaining feedback on a draft questionnaire from a number of people with different levels of knowledge about the topic of the survey. This can be a tremendously valuable input to survey design, particularly when there is a limited time to design and administer the survey. This type of feedback can give a quick insight into the different aspects of an effective questionnaire design, for example, to be sure that questions are measuring what was intended to be measured and that potential technical or vocabulary issues are identified. An initial recommendation here would be to ask at least 3 to 5 people who are not experts in the topic or who were not engaged in designing in the questionnaire. If there is additional time and resources, the next step would be to conduct 15 to 20 think-aloud interviews. This was what was done in relation to the 2017 SPIRE survey and underpins the recommendations that follow.

There are multiple factors to consider when creating an effective impact survey and structuring good survey questions: what type of questions to write (e.g., open vs. close-ended); how to word the question appropriately; what response options to provide; what should be the visual layout; whether to provide additional instructions. In the next section, we discuss and illustrate the six most common issues identified from a series of in-depth interviews as the most likely to jeopardise effective impact reporting.

3.2. Specific Issues

3.2.1. Inappropriate Language

On the surface, writing a question may seem like a simple action. However, making sure that appropriate language is used requires special attention. In our think aloud interviews we identified multiple points in the 2017 SPIRE questionnaire where project coordinators did not understand the meaning of a particular concept or question. For example: “*Does the project develop high-skilled jobs with new profiles? yes/no*” (Q29). Yet the decision to write such a question right away raises a whole set of potential issues. There is a concern that different formats might produce different answers for the same question. Also, is the respondent familiar with the concept of *high-skilled jobs with new profiles* and is the reader interpreting the questions as it was intended?

Having a professionally constructed questionnaire can be an effective way to establish credibility with respondents. However, there is also a danger that this might lead to the use of complex and technical words and phrases that respondents might struggle with [49]. A recommended rule of thumb in this situation is that when a word exceeds six or seven letters, a shorter and more easily understood word could potentially be substituted [40]. Another common tendency is to use abbreviations or specific jargon familiar to the survey designers such as the European Commission or government agencies; but this might create additional confusion for the respondents—for example, abbreviations such as TRL (Technology Readiness Level). The recommended practice here is always on the first occurrence to use the full name and only subsequently to introduce abbreviations. For later uses of such words or phrases it may be appropriate to replace complex and specialised words with abbreviations but there can still be instances where it is not necessary or even wise. For example, from the think aloud interviews about the 2017 SPIRE survey, GTG (Gate-to-Gate), CTG (Cradle-to-Gate) and CtGv (Cradle-to-Grave) even after having been initially defined and mentioned, were still confusing for respondents later in the survey table. In this case, each of the options sounds relatively similar and abbreviations look fairly similar as well, which led to confusion when trying to provide an answer.

Writing plainly means writing to be understood, using familiar language in a logical presentation. Plain language does not mean dumbing down the content of the questions but rather writing in a way that non-experts in the field can understand. Survey questions might be easier to comprehend when they use shorter everyday words (such as “bleeding”) rather than official terminology (such as “haemorrhage”) [50]. This can be a particular issue with sustainability topics where respondents can often be an audience with diverse backgrounds and different levels of experience. For example, some people may interpret the “greenhouse effect” as causing local weather changes rather than overall climate change [7]. This was often the case in our interviews, where project coordinators misinterpreted the meaning of concepts from the questionnaire because of their diverse educational background in terms of academic disciplines studied. In designing an effective impact measurement tool, it is important that these variations do not harm the data gathered.

A general guideline for crafting good questions is to use specific and concrete words to explain the concept clearly. A good example from the questionnaire is: “*How many SMEs are involved in the project as partners?*” (Q3). All respondents clearly understood the question and did not have a problem with providing a numerical answer. It is crucial to make sure that the concepts in the questionnaire are clearly defined and communicated in order to diminish the amount of interpretation required from respondents. To avoid overestimating respondents’ level of knowledge and vocabulary, it is

recommended to run a pilot-test to identify if there are any difficulties with understanding the content. This recommendation will be elaborated further in the final recommendation (see Section 4).

Sometimes using simple wording can prompt respondents to give different answers. One example of a project that benefited from think aloud interviews was conducted by an interdisciplinary team of economists and a psychologist [47]. This project aimed to improve the inflation expectations questions that are commonly asked on national consumer surveys. The existing survey had been measuring Americans' inflation expectations for more than 50 years [51], by asking "During the next 12 months, do you think that prices in general will go up or go down or stay where they are now?" followed by the response options "Go up," "Stay the same," and "Go down." Those who responded "go up" or "go down" were then asked to give a specific percentage. The issue was that even in times of relatively stable inflation, survey results showed large disagreements between respondents [52,53]. The interviews revealed that, while some individuals recognise the "prices in general" wording as referring to "inflation" others thought about their *personal* price experiences. The follow-up surveys showed that thinking more about personal price experiences when answering questions about expectations for "prices in general" was associated with giving much higher responses [47]. The recommendation in this study was to ask directly about expectations for inflation. Thus, while simpler wording (such as "prices in general") is generally recommended in the survey design literature, it should be noted that sometimes more technical but specific terms (such as "inflation") do a better job at communicating the question designers' intent and reducing respondents' disagreements about what the question means.

One strategy to check survey readability levels is to use language tools. For example, the Flesch-Kincaid Readability formula is commonly used in survey design and risk communication to measure the readability of a text [54]. The general recommendation here is to write at a level that 11–12 year olds can understand. Other sources might include websites like Read-able.com or Readability-score.com that can provide support quickly to estimate the content's reading level by combining results from different reading level indices. Another useful tool is Hemingway Editor that points out complicated sentences, common errors, passive voice and adverbs and reports reading scores.

3.2.2. Two-in-One Questions and Related Biases

The recommendation to ask one question at a time might seem obvious, yet it is surprising how often questionnaires ask questions that contains two components about which respondents might think or feel differently [40]. An example from our interviews is as follows. Respondents were asked to report "if they expect that their project will have impact in terms of job preservation or job creation?" (Q30). Responses options were "Yes and please give estimates" and "No." Respondents found this question particularly difficult to provide an estimate for and also were uncertain for what to provide an estimate if they had both job preservation and job creation to report. As this question is written, it poses a problem for respondents who want to reply only for job creation or just job preservation because these are two related but distinctively different concepts. Questions framed in this way also pose a problem when it comes to interpretation because survey users would not know to which component respondents were referring when they marked "yes" or "no." One possible solution to this problem is to separate both concepts by asking two questions instead. For example: "Do you expect your project to have impacts in terms of job preservation?" "yes" and "no" and "Do you expect your project to have impacts in terms of job creation?" "yes" and "no."

Survey questions often consist of multiple parts that work together to produce a high-quality output about the topic of interest. If one part of the question fails or provides a conflicting message with another part, it can undermine the accuracy of the answers provided. Therefore, when creating good survey questions, it is important to understand how each element of the questions conveys meaning independently to the respondent as well as all combined. The wording of the question is one of the most essential parts of crafting a good question. In addition to the question wording

itself, additional instructions and definitions or examples that will help respondents to understand the meaning of the question better can also be considered.

Another issue can be asking a question that does not apply to every respondent. For example, in the 2017 SPIRE survey, respondents were asked *“are/will you meet resource efficiency targets of the proposal?”* (Q23) Response options were “yes,” “partially” and “no.” If “partially” or “no” was answered, respondents were asked to provide a brief explanation why. The issue with this question is that it assumes that every project has efficiency targets set as a measure of sustainability progress. Also, participants reported that they perceived that they “felt pressure” to answer positively to this question potentially revealing “positive response bias” [55] and that the answer closely depends on the stage of the project. The issue of tailoring the questions to the stage of the project will be discussed further in one of the following recommendations (see Section 3.2.5).

This type of question can be especially damaging when participants are required to enter an answer for every question before they are allowed to advance to the next one. The participant here is faced with two options if this question does not apply to them: (i) consciously enter incorrect or nonsense information or (ii) quit filling out the survey altogether. The analysis of survey data in our study confirmed such a pattern of responses. A good rule of thumb in the situation with this type of question is to first ask: *“Do you have an efficiency target in your project?”* Depending on their answer, this can then be followed up with another related question. This example illustrates the challenge of crafting good questions that every potential participant will interpret in the way intended and will be able to respond to accurately.

It might be tempting to reduce the number of words in questions by presenting only one side of an issue. For example: *“Are/will resource efficiency targets of the proposal be met? Yes/Partially/No”* (Q23). A question written in this way implicitly suggests that your project should meet resource efficiency targets. It underlines the “desirable” particularly because the continuation of these proposed options says: *“If partially or no, please provide brief explanations why?”* with an open-ended answer box. Framing the question in this way might create a bias where participants will favour one option over another or where there could be a perceived risk that, because of the language used to frame the question, answering partially or no could result in funding being withdrawn, thus prompting a false positive answer. In either case, this can lead to misinterpretation of data collected and overoptimistically reporting impact of the projects.

3.2.3. Over Hypothetical Questions

It might be easier to get more precise answers for some types of survey question than others. For example, by asking the question: *“How old are you?”* we are expecting that participants should not have any major difficulties in providing an answer, because it is a fairly common question and people are usually well-aware of their age. This type of question is usually accurately reported, similarly to other factual or demographic questions. However, asking a question that is more abstract, less familiar or more hypothetical might well require additional effort from respondents. Frequently survey designers or administrators want their respondents to answer questions that require estimates or guesses. For example: *“Please, describe the scenario of potential deployment for the exploitable results you assume by 2030? Please indicate expected market size and market share”* (Q10). Sometimes to provide an answer for such questions might be especially difficult, if not impossible. In this example above, the projects they are reporting might not have a business or market plan as far as the year 2030 or they may not have conducted such an analysis yet. To avoid this tendency, survey designers should provide more concrete and less hypothetical questions such as asking for the next 5 or 10 years. If asking about estimates, survey designers should also capture any associated uncertainty by asking about the confidence level respondents have in their responses. This will indicate the levels of uncertainty for each of those responses. The higher the uncertainty, the stronger the disclaimer when interpreting the reported information. This issue can be resolved by asking participants to provide their confidence judgments to capture the uncertainty of their estimates about the specific concept [56].

There is a danger here that this type of question can be easily influenced by the visual layout, the options offered, the wording of the questions or by reference points. Reference points can be used as a starting point to make judgments about different situations [57]. For example, providing a single or multiple reference points can serve as a simple guide and help participants to provide more accurate [58,59]. In addition, choosing an appropriate reference range, using definitions and examples, can also help respondents to provide more meaningful responses [60].

3.2.4. Inadequate Alternatives and Formats

There are two broad types of question format: open-ended and closed-ended questions. The open-ended question format provides a blank space or box where participants write their answers in their own words while the closed-ended question format provides participants with a list of created options from which they must select an answer to the question.

The open-ended question format enables participants freely to answer the question and it is preferred when the goal of the question is to receive detailed information about the topic. The most common type of open-ended question is the descriptive question, where participants provide in-depth information on the topic of the question. For example, *“Would you have further communication materials that can help us to better understand the impacts of the project (please, provide a link)?”* (Q5). The survey design literature suggests that by providing extra motivation to respond such as *“Your answer to this question is very important for understating this issue”* influences response length by between 5 and 15 words, increases the response rate from 12 to 20 percentage points and expands the time participants spend on providing the answer [40].

Another question type that was prevalent in the 2017 SPIRE questionnaire was the open-ended question for numerical responses such as frequency, amounts, percentages and other numerical values. The purpose of such a question format is to ask participants to enter a single number or amount into the answering box. For example, *“Please, indicate markets concerned by the functions of the exploitable results described in Table 7. Indicate current market size (in EUR million) in the EU _____ and outside the EU _____”* (Q9). The recommendation with numerical responses is always to ask for specific units of measurement to avoid nonsense responses. Also, for both descriptive and numerical types, the answer spaces should be appropriately sized for the response because previous research has shown that if the boxes are too large participants are more likely to enter extra information [61]. This might be problematic if the question format is open-ended in a large table that requires multiple numerical responses. The overall disadvantage of open-ended questions is that participants might get discouraged with this type of format and skip a question and if they do respond it might be more complex to analyse their responses.

The closed-ended question format is usually used when the survey designer has created a set of answer choices for participants. The choices can be presented on nominal (participants are asked to select a set of choices with no natural order implied) or ordinal (participants are provided with a set of choices with an order and they have to decide where their response fits in the given ranking) scales. Having an overview of different question formats can help survey designers craft an effective question that truly captures what is intended. For example, *“Which sectors of the process industry are involved in the project?”* (Q2) was presented as an open-ended question. However, changing the format of this question from an open-ended to closed-ended question would have saved time and effort for the respondents. In addition to being subject to error, the coding of open-ended responses can also be very time consuming, especially when surveys have many respondents [47]. In this case developing a list of answer categories that includes all reasonable possible answers would be appropriate (such as cement, ceramics and others). What is important when developing a list of answer categories is that they are mutually exclusive and (assuming an ‘other’ category) collectively exhaustive. For example, *“Will the project lead to launching any of the following into the market? (i) New product (good, technology, service); (ii) new process; (iii) new method”* (Q6). Participants struggled to provide a clear answer since in some cases they could not differentiate between the options (quote from the interviews: *“If product is a*

technology, then it is also, a process. There should be more explanations of this concept . . . ”). Also, answer categories in this case did not provide additional clarification or guidance about how they should be interpreted. Failing to choose the appropriate answer categories can lead to potential biasing responses. For example, it might lead to participants choosing one option more (or less) frequently [60].

In the questionnaire we analysed in our study there were a few combinations with partially-closed question formats such as: “Do you foresee any policy barriers for the deployment of the project’s exploitable results in the involved SPIRE sectors and in other industrial sectors? Yes/No. If yes, please specify: _____” (Q18). This format can be useful. It represents a hybrid open-end/closed-ended question format allowing participants who do not fit into the provided response category to add another one. Alternatively, if a particular set of choices is important for survey designers, then it is possible to allow an option to add more content. In online surveys, a survey flow with these types of follow-up question where a second question depends on the answer to the first is much easier to implement.

A potential issue with this type of wording is that the survey designer is assuming that the respondent will definitely know the answer. However, the results from our think-aloud interviews showed that not all participants were familiar with every question and did not always know the answer. With just two alternatives, yes and no, when analysing the results, it might be interpreted as there being no real policy barriers but the reality might be that participants were simply unsure what to answer or did not know the answer. Adding an: “I don’t know” option might serve to supply data that can be more reliably interpreted.

Apart from choosing words and forming clear questions, visual design and layout of the questions in a survey influence participants as well. Survey designers need to take into account all the elements of an effective questionnaire. There are three key guidelines that should be applied when it comes to visual presentation of survey questions: (i) make sure that the words and visual elements that make up the question send consistent messages [62]; (ii) use visual aspects of presentation to emphasise questions that are important; (iii) organise each question and additional instructions in a way that reduces the need to reread the question to increase overall comprehension.

3.2.5. Questions Not Tailored to the Stage of the Project

When thinking how to formulate questions for a questionnaire, there are three golden rules to follow [40]. First, choose the first question carefully. The rule of thumb is to pick a question that should apply to everyone and is easy to comprehend and answer. Secondly, place at the beginning the questions that are the most important for the survey designers or administrators. Thirdly, group questions that are related by topic because switching between topics might confuse respondents. In the 2017 SPIRE questionnaire, four parts based on impacts such as environmental, business, socio-economic impact and project identification were used, plus a Success Stories part. This meant that all the questions with similar topics were grouped together, making it easier for respondents to follow.

Another important decision when designing an effective questionnaire is how many questions to present on each screen or page. On the one hand, presenting all the questions on the same page can give respondents an overview of an entire questionnaire so they can make more informed decisions about whether and how to complete the survey [63]. However, this format has certain drawbacks. It can increase the chances that, while scrolling down, respondents might miss a question. It might discourage respondents to commit and complete the entire survey, especially if it is a longer questionnaire. Also, it limits the use of different interactive features of web surveys, such as asking follow-up questions based on previous answers. On the other hand, presenting each of the question/s on its own page allows participants to focus on one question at time. In contrast to presenting all at once, different online survey features can be implemented. If survey designers want to keep participants informed about the survey length, they can insert a progress indicator that can show the percentage of completeness at any given stage. The disadvantages of such a design are that it often takes longer to complete, it requires

additional clicks and it can be problematic if participants need to remember what they have answered on previous questions, which was the case with the main table in the 2017 SPIRE questionnaire [64].

If policy-makers want to design a user-friendly questionnaire, they should avoid making responses to questions compulsory unless necessary. In some cases, requesting an answer to one or more questions can be crucial to the goal of the survey and can save time and expense. However, for most questionnaires, explicitly requesting an answer for each of the questions can have a harmful effect on respondents' motivation, quality of responses and the likelihood that participants will complete the full survey [65]. This was reported from the project coordinators in our study, that when they were forced to provide an answer to a question where they did not know what to write, they would often provide nonsense answers such as 0.9999 or simply write not applicable. In addition, the majority of institutional review boards or ethical committees would strongly recommend that participants should be allowed to skip questions they prefer not to answer [66].

If policy-makers decide to require responses for all questions, they should provide a "not applicable" option for the respondents. One of the biggest issues with the 2017 SPIRE questionnaire was that it was not customized and tailored to the stage of the project. For example, "*Please provide one feature (a success story) to promote your project? Please indicate the feature shortly (2–3 lines) and suggest a person to contact for further information (if needed)*" (Q4). Success stories were requested to be reported in the questionnaire even when a project was in its initial phase, thus respondents were not necessarily able to provide an example of a success story.

The survey design literature suggests that shorter and respondent-friendly questionnaires can improve response rates [67,68]. Thus, excluding questions that are not applicable to the stage of the project and designing questionnaires to minimize the burden can improve the way that respondents provide their answers. In addition, if there is more than one questionnaire for respondents to respond to over the years of a project, it is best to try to make sure that they all follow the same design principles and that the repeating questions are asked in a similar or identical way [69] and/or are tentatively pre-populated. Furthermore, if the survey is being conducted on an annual basis and will be used to compare the results in some way, the general recommendation is to avoid or only to apply minimal changes to the survey. However, if it is determined that a specific question might be problematic for various methodological reasons such as validity of the question, that question should be re-worded in a careful manner to accurately address the identified issue.

3.2.6. Lack of Additional Guidance

It is valuable to create interesting and informative welcome and closing screens that will appeal to the participants. These screens are exceptionally important because it is the first thing that participants will be exposed to. It is also important for setting the right tone by providing a description of the survey and instructions for how to proceed. In addition, develop a screen format that emphasises the respondents rather than the survey designer or administrator. This is particularly important when it is an experienced person who is gathering and reporting those metrics and receiving the questionnaire. If a project coordinator does not understand specific language in a question, it is possible to help them by providing additional explanations and definitions. In other words, we would recommend creating a document with guidelines that contains additional explanations and instructions for specific questions to avoid misinterpretation.

Sharing in advance information about the questionnaire with the participants including how the results will be used to benefit them and others can increase participation rates [69]. For example, sending information leaflets or other communication materials that provide insights into the importance of the survey and why the survey is being conducted might positively encourage people to respond.

Furthermore, questionnaires that explicitly ask for respondents' advice and feedback, such as including the phrase "*it would really help us out,*" showing that their input is important for overall project evaluation increases response rates by 19%, according to Mowen and Cialdini [70]. In addition, phrases of verbal appreciation such as "*we appreciate your help*" increase the likelihood of people

responding to the questionnaires [71]. Reporting SPIRE success stories helps to publicise information from projects so instead of just saying “*it would really help us out,*” we could give the message that “this will allow us to help you.”

Ensuring that respondents are motivated to respond to each of the questions in a meaningful way should always be a major concern of every survey designer or administrator. Especially if the results of the survey will be used to write a report that will be shared with the public this is important, because without proper motivation or incentive, participants may ignore the instructions, skip the questions, provide incomplete answers or fail to complete the survey in total. Another cost-effective suggestion to inform respondents is to organise webinars or create online videos about how to approach and complete the questionnaire. This can allow some misinterpretations to be resolved on the spot (in the case of online web events).

4. Conclusions and Recommendations

To improve impact reporting tools that serve industry on a day-to-day basis requires a pragmatic design that will help project coordinators or any person in charge of those activities to make better and more informed decisions about their sustainability practices. The question of whether to report sustainability metrics is no longer an issue [13] but rather what and how should they be reported to accurately demonstrate the impact [15,16].

We have conducted think aloud interviews exploiting the mental models approach to communications design [34] in which project coordinators in the SPIRE network read through the questionnaire materials, thought out loud while providing answers for the questions and gave comprehensive feedback about their experience. Our findings suggest that participants who completed the SPIRE questionnaire took it seriously and tried to share as much information as they had available about the project. However, misunderstandings were reported across the questionnaire, related to its design. Specifically, the results show that factors such as ambiguous terminology, two-in-one questions, the stage of the project, over-hypothetical estimates and inexperience of the project coordinator can negatively influence the data collection process and jeopardise the overall validity of the findings. All these identified issues can potentially jeopardise the accuracy of the data, validity of the questionnaire and monitoring efforts of the project. Six key guidelines have been formulated as a result of the think-aloud interview transcripts analysis (see Table 2 below).

Firstly, having a professionally formalised questionnaire is an effective way to establish credibility with respondents. However, there is a danger to be guarded against, that this can lead to use of complex and technical words and phrases that they might struggle with. Secondly, think about the survey question set as a whole to ensure they are coherent, consistent and relevant to all respondents (or, when not, that it is clear how to respond). Avoid two-in-one questions or wording that may implicitly bias the response. Thirdly, asking questions that are more abstract, less familiar or more hypothetical can make the respondent’s task unnecessarily hard. It is recommended to ask more concrete questions, such as asking for shorter future time predictions and also, in appropriate situations, to ask respondents to provide their confidence judgments to capture the uncertainty of their estimates about specific concepts.

Fourthly, survey designers should choose question format and visual design carefully, bearing in mind the type of information they are hoping to receive—be clear about units of measurement; think carefully about the choice between open, closed and partially closed questions. Fifthly, think about the order and grouping in which questions are presented and the logic from the respondent’s viewpoint. Where projects may be at very different stages of completion, make sure that all questions remain meaningful independent of project stage or that it is clear how to respond when a question is not appropriate to the stage of the project concerned. Finally, create a document or other guidelines that contain additional explanations and instructions for specific questions to avoid misinterpretation. Think also about what might engage respondents and motivate them to contribute fully and thoughtfully.

Table 2. Key findings and best practices for impact reporting.

Issues Identified:	Suggested Solutions:
1. Use of inappropriate language	<ol style="list-style-type: none"> 1. Use appropriate terminology and rewrite the existing question that were identified as confusing. 2. A recommended rule of thumb when a word exceeds six or seven letters, a shorter and more easily understood word could probably be substitute. 3. The recommended practice with abbreviations is always on the first occurrence to use the full name and only subsequently to introduce abbreviations 4. In some cases, more technical but specific terms do a better job at communicating the question designers' intent and reducing respondents' disagreements about what the question means. 5. Use the Flesch-Kincaid Readability formula to measures the readability of a text 6. Run a short Pilot test to identify if there are any difficulties with understanding the content and the meaning of specific concepts.
2. Two-in-one questions and related biases	<ol style="list-style-type: none"> 1. Make sure that you only ask one question at time. 2. Frame questions that does not implicitly suggests desirable answers. 3. Allow respondents to progress to different part of the questionnaire without demanding to populate all previous questions.
3. Focus on over hypothetical questions	<ol style="list-style-type: none"> 1. Ask for shorter time periods when requiring respondents to provide numerical estimates. 2. Include reference points or previous examples to serve as a simple guide for respondents
4. Use inadequate alternatives and formats	<ol style="list-style-type: none"> 1. By changing the question format from open-ended to multiple choice can decrease the response time. 2. When creating a question with multiple alternatives, make sure developing a list of answer categories that include all reasonable possible answers that are mutually exclusive. 3. Offer options that are distinctive and additional examples for each of the choices. 4. With numerical responses ask for specific units of measurement to avoid nonsense responses 5. Add "I don't know option" in some cases to avoid assumption that respondents are familiar with every question asked.
5. Questions not tailored to the stage of the project	<ol style="list-style-type: none"> 1. Tailor the question to the stage of the project. 2. Put up-front the questions that ate most important for the survey 3. Include a progress indicator while completing a survey 4. Add "Not applicable option" in case when questions do not apply to every respondent
6. Lack of additional guidance	<ol style="list-style-type: none"> 1. Provide more guidelines on how to attempt to answer this question for specific project. 2. At the end of the survey, ask respondents for feedback that can help to improve the future design. 3. For new projects develop a webinar or online a video that can be used to inform respondents.

As with all questionnaires, pre-testing is one of the key elements for delivering an effective survey design. Many survey designers and administrators fail to pre-test their questionnaires or ask colleagues who are experts in this domain or have been involved in creating the survey. The recommendation here is to ask at least 3 to 5 people who are not experts in the topic or were engaged in designing the questionnaire. If there is additional time and resources, the next step in survey design would be to conduct 15 to 20 think-aloud interviews of the complete questionnaire in order to identify wording, question order, visual design and navigation related difficulties [72]. This type of interview is the best method to test questions and questionnaires. Analysis using this approach was of the basis for evaluating the questionnaire about sustainability analysed in this paper.

Developing survey protocols to make sure the process of evaluation is systematic is recommended. In situations where the results will directly inform specific governments, organisational policy or a programme, it is advised that an experimental evaluation of questionnaire components is conducted. Experimental methods, in addition to qualitative evidence from the interviews, will provide quantitative estimates of the effects of proposed changes in questionnaires on much larger samples. The mental model approach represents the most powerful basis for evaluation of survey questions and questionnaires. If there is enough time and resources, the survey administrator should conduct a small pilot study with a sample of the population in order to evaluate the survey. The aim of such a study is to determine whether the proposed survey and procedures are appropriate for the final survey implementation. Pilot studies can give a good insight on how the study conducting a questionnaire will look like and if initial collected data make sense.

A common tendency for survey designers is to overestimate the vocabulary of respondents as well as their level of knowledge. The best way to overcome this issue is to pilot test its questions with members of the population of interest to identify potential problems. Also, it can be useful in large-scale projects to randomise the order of response options within a question or questions to avoid order effects due to memory or cognitive limitations. In surveys that require a longer time to fill out, participants may become more confused as they try to provide an answer and have more information to keep track of.

In addition to testing questions and survey design, there are additional steps to consider in the long run. Building trust that responses provided will inform future initiatives and enable the organisation to deliver better programmes and outcomes is important for the respondents providing the answers; they may engage again with the initiative concerned in the future if they had a positive relationship. Further, ensuring security of information, that the data collected in this questionnaire will be highly safeguarded is another key driver of trust and continuing engagement.

It is the argument of this paper that future impact reporting should follow these recommendations to improve existing questionnaires and when developing new ones. However, these suggestions go beyond just impact reporting in process industries and can be applicable to other types of industry that have similar reporting practices such as transportation or agriculture. Having well-designed questionnaires that accurately portray the impact of a project is important both at the project level and for all sustainability-related decisions. If sustainability targets are not being reached or their monitoring is not calibrated well enough, we might well fail to invest and ensure that real-world impact is achieved. If we do not design user-friendly interfaces, it might be that the person responsible for reporting such data would get frustrated and demotivated and fail to report data required for valid evaluation of a project. Also, in order to ensure that impact reporting is accurate and unbiased, reporting on the impact of sustainable actions should be related to the external impacts and stakeholder interpretation and not only with project managers. Combining these elements of careful consideration with what metrics are included in any questionnaire together with informed and thoughtful decisions about how the metrics are collected and recorded is crucial to successful impact reporting.

Author Contributions: Conceptualization: V.L., R.H., A.P. and A.P.; Data curation: V.L. and A.P.; Formal analysis: V.L.; Funding acquisition: R.H. and A.P.; Investigation: V.L., R.H., A.P. and A.P.; Methodology: V.L. and A.P.;

Project administration: R.H., A.P. and A.P.; Resources: A.P.; Supervision: R.H. and A.P.; Validation: R.H. and A.P.; Writing—original draft: V.L. and R.H.; Writing—review & editing: R.H., A.P. and A.P.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 767412.

Acknowledgments: We would like to thank Wändi Bruine de Bruin, Greet Van Eetvelde, Mark Talford, Gareth Jenkins, Nicolas Segebarth and Vladimir Cid-Bourié for their time and feedback.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. International Energy Agency. World Energy Outlook 2015. 2015. Available online: <https://www.iea.org/publications/freepublications/publication/WEO2015.pdf> (accessed on 3 April 2018).
2. Butler, R. A New Leaf in the Rainforest: Longtime Villain Vows Reform. *Yale Environment* 360, 10 March 2014.
3. Woods, C.L. From an “outside group” to a catalyst for corporate change: Examining activists’ legitimacy strategies. *Publ. Relat. Rev.* **2019**. In Press. [CrossRef]
4. Orji, I.J. Examining barriers to organizational change for sustainability and drivers of sustainable performance in the metal manufacturing industry. *Resour. Conservat. Recycl.* **2019**, *140*, 102–114. [CrossRef]
5. Kolk, A. Trends in sustainability reporting by the Fortune Global 250. *Bus. Strateg. Environ.* **2003**, *12*, 279–291. [CrossRef]
6. Masud, M.; Hossain, M.; Kim, J. Is Green Regulation Effective or a Failure: Comparative Analysis between Bangladesh Bank (BB) Green Guidelines and Global Reporting Initiative Guidelines. *Sustainability* **2018**, *10*, 1267. [CrossRef]
7. Bostrom, A.; Morgan, M.G.; Fischhoff, B.; Read, D. What Do People Know About Global Climate Change? 1. Mental Models. *Risk Anal.* **1994**, *14*, 959–970. [CrossRef]
8. Wilmshurst, T.D.; Frost, G.R. Corporate environmental reporting. *Account. Audit. Account. J.* **2000**, *13*, 10–26. [CrossRef]
9. Worrell, E.; Allwood, J.; Gutowski, T. The Role of Material Efficiency in Environmental Stewardship. *Ann. Rev. Environ. Resour.* **2016**, *41*, 575–598. [CrossRef]
10. Peace, A.; Ramirez, A.; Broeren, M.L.M.; Coleman, N.; Chaput, I.; Rydberg, T.; Sauvion, G.N.-N. Everyday Industry—Pragmatic approaches for integrating sustainability into industry decision making. *Sustain. Prod. Consum.* **2018**, *13*, 93–101. [CrossRef]
11. Tello, P.; Weerdmeester, R. *Sustainable Process Industry through Resource and Energy Efficiency*; SPIRE: Brussels, Belgium, 2013.
12. Breckell, P.; Harrison, K.; Robert, N. *Impact Reporting in the UK Charity Sector*; CFDG/Cass Business School: London, UK, 2010.
13. White, A.L.; Zinkl, D. *Green Metrics: A Global Status Report on Standardized Corporate Environmental Reporting*; Ceres Annual Conference: Boston, MA, USA, 1998.
14. Manes-Rossi, F.; Tiron-Tudor, A.; Nicolò, A.; Zanellato, G. Ensuring More Sustainable Reporting in Europe Using Non-Financial Disclosure—De Facto and De Jure Evidence. *Sustainability* **2018**, *10*, 1162. [CrossRef]
15. Batterham, R.J. Sustainability—The next chapter. *Chem. Eng. Sci.* **2006**, *61*, 4188–4193. [CrossRef]
16. Jose, A.; Lee, S.M. Environmental Reporting of Global Corporations: A Content Analysis based on Website Disclosures. *J. Bus. Ethics* **2006**, *72*, 307–321. [CrossRef]
17. Scott Marshall, R.; Brown, D. Corporate environmental reporting: What’s in a metric? *Bus. Strateg. Environ.* **2003**, *12*, 87–106. [CrossRef]
18. Bansal, P.; Roth, K. Why Companies Go Green: A Model of Ecological Responsiveness. *Acad. Manag. J.* **2000**, *43*, 717–736.
19. O’Donovan, G. *Legitimacy Theory and Corporate Environmental Disclosure: Some Case Study Evidence*; Accounting Association of Australia and New Zealand Annual Conference: Hobart, Australia, 1997.

20. Miles, M.P.; Covin, J.G. Environmental Marketing: A Source of Reputational, Competitive, and Financial Advantage. *J. Bus. Ethics* **2000**, *23*, 299–311. [[CrossRef](#)]
21. Guthrie, J.; Abeysekera, I. Content analysis of social, environmental reporting: What is new? *J. Hum. Resour. Cost. Account.* **2006**, *10*, 114–126. [[CrossRef](#)]
22. KPMG. *KPMG International Survey of Environmental Reporting*; KPMG Environmental Consulting: DeMeern, The Netherlands, 2017.
23. Epstein, M.J. *Making Sustainability Work: Best Practices in Managing and Measuring Corporate Social, Environmental and Economic Impacts*; Routledge: London, UK, 2018.
24. STYLE. STYLE (“Sustainability Toolkit for Easy Lifecycle Evaluation”) is a European Project that Aimed to Analyse the Existing Situation with Available Sustainability Evaluation Tools and Enable Project Teams to Make Better Decisions about Their Options. 2017. Available online: <https://www.spire2030.eu/sites/default/files/users/user221/STYLE/STYLE-Roadmap.pdf> (accessed on 2 April 2018).
25. MEASURE. Multi-Criteria Decision Analysis Tools Training Module. 2017. Available online: <https://www.spire2030.eu/sites/default/files/project/measure/uploads/Modules/Mediaroom/measure-training-module-mcda.pptx> (accessed on 2 April 2018).
26. SAMT. SAMT Report—Deliverable D1.1. 2017. Available online: https://www.spire2030.eu/sites/default/files/users/user355/SAMT_D.1.1_final_updatedlinks_Dec2016.pdf (accessed on 3 April 2018).
27. SPRING. SPRING Project. 2019. Available online: <https://www.spire2030.eu/spring> (accessed on 30 October 2018).
28. Yakovleva, N.; Kotilainen, J.; Toivakka, M. Reflections on the opportunities for mining companies to contribute to the United Nations Sustainable Development Goals in sub-Saharan Africa. *Extr. Ind. Soc.* **2017**, *4*, 426–433. [[CrossRef](#)]
29. European Environment Agency. *Digest of EEA indicators—Technical Report No 8/2014*; European Environment Agency: Copenhagen, Denmark, 2014.
30. Smeets, E.; Weterings, R. *Environmental Indicators: Typology and Overview*; European Environment Agency: Copenhagen, Denmark, 1999.
31. Azapagic, A.; Millington, A.; Collett, A. A Methodology for Integrating Sustainability Considerations into Process Design. *Chem. Eng. Res. Des.* **2006**, *84*, 439–452. [[CrossRef](#)]
32. Tufvesson, L.M.; Tufvesson, P.; Woodley, J.M.; Börjesson, P. Life cycle assessment in green chemistry: Overview of key parameters and methodological concerns. *Int. J. Life Cycle Assess.* **2012**, *18*, 431–444. [[CrossRef](#)]
33. Tobin, G.; Bell, H. Efficient and effective? The 100-year flood in the communication and perception of flood risk. *Environ. Hazards* **2007**, *7*, 302–311.
34. Bruine de Bruin, W.; Bostrom, A. Assessing what to address in science communication. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14062–14068. [[CrossRef](#)] [[PubMed](#)]
35. Morgan, M.G.; Fischhoff, B.; Bostrom, A.; Atman, C.J. *Risk Communication*; Cambridge University Press: Cambridge, UK, 2001.
36. Bruine de Bruin, W.; van der Klaauw, W.; van Rooij, M.; Teppa, F.; de Vos, K. Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise. *J. Econ. Psychol.* **2017**, *59*, 45–58. [[CrossRef](#)]
37. Jaspers, M.W.; Steen, T.; Van Den Bos, C.; Geenen, M. The think aloud method: A guide to user interface design. *Int. J. Med. Inform.* **2004**, *73*, 781–795. [[CrossRef](#)]
38. Cotton, D.; Gresty, K. Reflecting on the think-aloud method for evaluating e-learning. *Br. J. Educ. Technol.* **2006**, *37*, 45–54. [[CrossRef](#)]
39. Bradburn, N. *Question-Wording Effects in Surveys*; Jossey-Bass: San Francisco, CA, USA, 1982.
40. Dillman, D.A.; Smyth, J.D.; Christian, L.M. Internet, phone, mail, and mixed-mode surveys. In *The Tailored Design Method*; Wiley: Hoboken, NJ, USA, 2014.
41. Schwarz, N. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*; Erlbaum: Hillsdale, NJ, USA, 1996.
42. Schwarz, N. Self-reports: How the questions shape the answers. *Am. Psychol.* **1999**, *54*, 93–105. [[CrossRef](#)]
43. Willis, G. *Cognitive Interviewing in Survey Design: State of the Science and Future Directions*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 103–107.
44. Peterson, C.H.; Peterson, N.A.; Powell, K.G. Cognitive Interviewing for Item Development: Validity Evidence Based on Content and Response Processes. *Meas. Eval. Counsel. Dev.* **2017**, *50*, 217–223. [[CrossRef](#)]

45. Murphy, M.; Hollinghurst, S.; Salisbury, C. Qualitative assessment of the primary care outcomes questionnaire: A cognitive interview study. *BMC Health Serv. Res.* **2018**, *18*, 79. [[CrossRef](#)] [[PubMed](#)]
46. Sooniste, T.; Granhag, P.A.; Strömwall, L.A.; Vrij, A. Statements about true and false intentions: Using the Cognitive Interview to magnify the differences. *Scand. J. Psychol.* **2015**, *56*, 371–378. [[CrossRef](#)] [[PubMed](#)]
47. Bruine de Bruin, W.; Baldassi, M.; Figner, B.; Fischhoff, B.; Fleishman, L.; Hardisty, D.; Li, Y. Framing effects in surveys: How respondents make sense of the questions we ask. In *Perspectives on Framing*; NY: Taylor and Francis Group; Psychology Press: New York, NY, USA, 2011; pp. 303–325.
48. Presser, S.; Couper, M.P.; Lessler, J.T.; Martin, E.; Martin, J.; Rothgeb, J.M.; Singer, E. Methods for Testing and Evaluating Survey Questions. *Public Opin. Q.* **2004**, *68*, 109–130. [[CrossRef](#)]
49. Rothman, R.L.; Housam, R.; Weiss, H.; Davis, D.; Gregory, R.; Gebretsadik, T.; Shintani, A.; Elasy, T.A. Patient Understanding of Food Labels. *Am. J. Prev. Med.* **2006**, *31*, 391–398. [[CrossRef](#)]
50. Lerner, E.B.; Jehle, D.V.K.; Janicke, D.M.; Moscati, R.M. Medical communication: Do our patients understand? *Am. J. Emerg. Med.* **2000**, *18*, 764–766. [[CrossRef](#)] [[PubMed](#)]
51. Curtin, R. *Inflation Expectations: Theoretical Models and Empirical Tests*; National Bank of Poland Workshop on The Role of Inflation Expectations in; National Bank of Poland Workshop: Warsaw, Poland, 2006.
52. Bryan, M.F.; Venkatu, G. *The Demographics of Inflation Opinion Surveys*; Economic Commentary 2001, No. 2; Federal Reserve Bank of Cleveland: Cleveland, OH, USA, 2001.
53. Ranyard, R.; Del Missier, F.; Bonini, N.; Duxbury, D.; Summers, B. Perceptions and expectations of price changes and inflation: A review and conceptual framework. *J. Econ. Psychol.* **2008**, *29*, 378–400. [[CrossRef](#)]
54. Flesch, R. A new readability yardstick. *J. Appl. Psychol.* **1948**, *32*, 221–233. [[CrossRef](#)]
55. Schaeffer, N.C.; Seltzer, J.A.; Klawitter, M. Estimating Nonresponse and Response Bias. *Sociol. Methods Res.* **1991**, *20*, 30–59. [[CrossRef](#)]
56. Bruine de Bruin, W.; Downs, S.J.; Fischhoff, B.; Palmgren, C. Development and Evaluation of an HIV/AIDS Knowledge Measure for Adolescents Focusing on Misconceptions. *J. HIV/AIDS Prev. Child. Youth* **2007**, *8*, 35–57. [[CrossRef](#)]
57. Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **1974**, *185*, 1124–1131. [[CrossRef](#)] [[PubMed](#)]
58. Lesic, V. Understanding and Informing Consumers' Perceptions of Their Electricity Use. Ph.D. Thesis, University of Leeds, Leeds, UK, 2017.
59. Lesic, V.; Bruine de Bruin, W.; Davis, M.C.; Krishnamurti, T.; Azevedo, I.M.L. Consumers' perceptions of energy use and energy savings: A literature review. *Environ. Res. Lett.* **2018**, *13*, 033004. [[CrossRef](#)]
60. Schaeffer, N.C.; Presser, S. The science of asking questions. *Ann. Rev. Sociol.* **2003**, *29*, 65–88. [[CrossRef](#)]
61. Christian, L.M.; Dillman, D.A.; Smyth, J.D. Helping Respondents Get It Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys. *Public Opin. Q.* **2007**, *71*, 113–125. [[CrossRef](#)]
62. Smyth, J.D.; Dillman, D.A.; Christian, L.M.; Stern, M.J. Effects of Using Visual Design Principles to Group Response Options in Web Surveys. *Int. J. Internet Sci.* **2006**, *1*, 6–16.
63. Crawford, S.D.; Couper, M.P.; Lamias, M.J. Web Surveys: Perceptions of Burden. *Soc. Sci. Comput. Rev.* **2001**, *19*, 146–162. [[CrossRef](#)]
64. Manfreda, K.L.; Batagelj, Z.; Vehovar, V. Design of Web Survey Questionnaires: Three Basic Experiments. *J. Comput.-Mediat. Commun.* **2002**, *7*, JCMC731. [[CrossRef](#)]
65. Tourangeau, R.; Rips, L.J.; Rasinski, K. *The Psychology of Survey Response*; Cambridge University Press: Cambridge, UK, 2000.
66. De Vaus, D. Surveys. In *Social Research*; Routledge: London, UK, 2013.
67. Heberlein, T.A.; Baumgartner, R. Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *Am. Sociol. Rev.* **1978**, *43*, 447–462. [[CrossRef](#)]
68. Dillman, D.A.; Sinclair, M.D.; Clark, J.R. Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys. *Public Opin. Q.* **1993**, *57*, 289–304. [[CrossRef](#)]
69. Groves, R.M.; Cialdini, R.B.; Couper, M.P. Understanding the Decision to Participate in a Survey. *Public Opin. Q.* **1992**, *56*, 475. [[CrossRef](#)]
70. Mowen, J.C.; Cialdini, R.B. On Implementing the Door-in-the-Face Compliance Technique in a Business Context. *J. Market. Res.* **1980**, *17*, 253–258. [[CrossRef](#)]

71. Blau, M.P. *Exchange and Power in Social Life*; Routledge: London, UK, 2017.
72. Sheatsley, P.B. Questionnaire Construction and Item Writing. In *Handbook of Survey Research*; Academic Press, Inc.: San Diego, CA, USA, 1983.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).