Year: 2020

Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow

Duboc, Leticia ; Penzenstadler, Birgit ; Porras, Jari ; Akinli Koçak, Sedef ; Betz, Stefanie ; Chitchyan, Ruzanna ; Leifler, Ola ; Seyff, Norbert ; Venters, Colin C

DOI: https://doi.org/10.1007/s00766-020-00336-y

Originally published at:
Duboc, Leticia; Penzenstadler, Birgit; Porras, Jari; Akinli Koçak, Sedef; Betz, Stefanie; Chitchyan, Ruzanna; Leifler, Ola; Seyff, Norbert; Venters, Colin C (2020). Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow. Requirements Engineering, 25(4):469-492.
DOI: https://doi.org/10.1007/s00766-020-00336-y

Cite as: Leticia Duboc, Birgit Penzenstadler, Jari Porras, Sedef Akinli Koçak, Stefanie Betz, Ruzanna Chitchyan, Ola Leifler, Norbert Seyff, Colin Venters (2020). Requirements Engineering for Sustainability: An Awareness Framework for Designing Software Systems for a Better Tomorrow. *Requirements Engineering* 25(4):469–492.

Requirements engineering for sustainability: an awareness framework for designing software systems for a better tomorrow

Leticia Duboc[1] · Birgit Penzenstadler[2,3] · Jari Porras[4] · Sedef Akinli Kocak[5] · Stefanie Betz[6] · Ruzanna Chitchyan[7] · Ola Leifler[8] · Norbert Seyff[9,10] · Colin C. Venters[11]

Abstract

Integrating novel software systems in our society, economy and environment can have far-reaching effects. As a result, software systems should be designed in such a way as to maintain or improve the sustainability of their intended socio-technical systems. However, a paradigm shift is required to raise awareness of software professionals on the potential sustainability effects of software systems. While Requirements Engineering is considered the key for driving this change, requirements engineers lack the knowledge, experience and methodological support for acting as facilitators for a broader discussion on sustainability effects. This paper presents a question-based framework for raising awareness of the potential effects of software systems on sustainability, as the first step towards enabling the required paradigm shift. An evaluation study of the framework was conducted with four groups of computer science students. The results of the study indicate that the framework is applicable to different types of systems and helps to facilitate discussions about the potential effects that software systems could have on sustainability.

Keywords Sustainability · Software · Socio-technical systems · Requirements engineering · Software engineering

The authors would like to acknowledge EPSRC Grant EP/R007373/1, Digitaldialog 21 funded by the Ministry for Science, Research and Art Baden-Württemberg, the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (grant 2017 SGR 966), and the funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 712949 (TECNIOspring PLUS) and from the Agency for Business Competitiveness of the Government of Catalonia.

✉ Colin C. Venters
c.venters@hud.ac.uk

Leticia Duboc
l.duboc@salle.url.edu

Birgit Penzenstadler
birgitp@chalmers.se

Jari Porras
jari.porras@lut.fi

Sedef Akinli Kocak
sedef.kocak@vectorinstitute.ai

Stefanie Betz
besi@hs-furtwangen.de

Ruzanna Chitchyan
r.chitchyan@bristol.ac.uk

Ola Leifler
ola.leifler@liu.se

Norbert Seyff
norbert.seyff@fhnw.ch

1 La Salle, Ramon Llull University, Barcelona, Spain

2 Chalmers University of Technology, Gothenburg, Sweden

3 Lappeenranta University of Technology, Lappeenranta, Finland

4 Lappeenranta-Lahti University of Technology LUT, Lappeenranta, Finland

5 Vector Institute, Toronto, Canada

6 Hochschule Furtwangen University, Furtwangen, Germany

7 University of Bristol, Bristol, UK

8 Linköpings Universitet, Linköping, Sweden

9 FHNW, Windisch, Switzerland

10 University of Zurich, Zurich, Switzerland

11 University of Huddersfield, Huddersfield, UK

Requirements Engineering (Springer)
DOI 10.1007/s00766-020-00336-y Preprint in final layout
Final publication available at https://link.springer.com/article/10.1007/s00766-020-00336-y

1 Introduction

Software underpins all aspects of societal life from commerce, communication, education, to energy, entertainment, finance, governance, health and defence. As a cornerstone of various socio-technical systems, software is also a key determinant of their sustainability, i.e. their capacity to endure [35]. Increasingly, it is advocated that sustainability requires simultaneous consideration of several interrelated dimensions (environmental, economic, individual, social and technical), which we refer to as the **five dimensions of sustainability** [36]. As a result, the sustainability of a socio-technical system should become a prime concern for the field of software and requirements engineering to address [6].

This paper advocates that, as major drivers of change within society, software systems must be designed to maintain the sustainability of the wider socio-technical system in which they are integrated [4, 18]. As such, requirements engineers have a degree of responsibility to support the discussion of the potential sustainability effects of the software across all dimensions of sustainability in order to account for potential (un-)desired consequences during the software's life cycle. Since the engineering of requirements is an inherently collaborative process (e.g. with product managers, direct and indirect stakeholders, etc.), we need a paradigm shift in requirements engineering practice, where all stakeholders involved in defining requirements take explicit responsibility for the sustainability effects of the technological solutions that they introduce into society. In this new paradigm, requirements engineers are facilitators of such discussions, helping stakeholders to translate their concerns into requirements. To achieve this, they can draw upon the lessons of tackling wicked problems from a holistic perspective advocated by the field of Systems Thinking [11] instead of the narrow computational thinking mindset of "solving a problem for a customer" [19]. The field of Requirements Engineering (RE) is the key to achieving this change [4]. We submit that such a change must start by raising awareness of the relationships between software systems and sustainability [41]. As such, our **key contributions** are:

1. A question-based "Sustainability Awareness Framework" (SuSAF) for raising awareness of the effects that a software system could have upon sustainability. Such *effects* are consequences of the production and prolonged use of the software (e.g. gentrification caused by Airbnb).
2. An evaluation of the proposed question-based framework with two instances of its application with students as part of a teaching curriculum.

This paper significantly extends the work presented in Duboc et al. [18] with three new research questions (RQ1, RQ4 and RQ5) . These questions look into the applicability of the framework to different types of systems and determine whether the framework's questions, its application process and supporting materials have helped to broaden the students' perspectives on sustainability effects. We have also detailed RQ6 with three sub-questions to gain further insights into access to relevant stakeholders and the usefulness and understandability of the framework. In order to answer these new research questions, this paper also significantly extends the analysis of the data set. Finally, we also expand the discussions on the design process used to create this framework, as well as related work.

The paper is organised as follows. Section 2 defines the concept of sustainability and a number of related terms. Section 3 summarises related research work, while Sect. 4 describes the SuSAF framework. The study design is outlined in Sect. 5, and the results are presented in Sect. 6. Section 7 discusses the lessons learned based on instructors' overall reflections and observations. Finally, the paper concludes in Sect. 8 by identifying some open issues and research challenges and highlights areas of future work.

2 Sustainability

Modern society's reliance on software systems has resulted in the emergence of sustainability as a growing area of interest in the field of software and requirements engineering [43]. In the context of this paper, sustainability is defined as the capacity of a socio-technical system to endure [4].

Two important and closely related concepts which extend the basic definition of sustainability are *sustainable use* and *sustainable development*. Hilty and Aebischer [23] define sustainable use of a system **S** with regard to a function **F** and a time horizon **T**, which in essence means to "use **S** in a way that does not compromise its ability to fulfil **F** for a period of **T**". This framing of sustainability aims to explicitly link the idea of use to the adjective *sustainable*, i.e. the ability to continue over a period of time. The Brundtland Commission defined sustainable development [8] as development that *meets the needs of the present without compromising the ability of future generations to meet their own needs*. The word "need" is central to this definition and includes a dimension of time, present and future, as well as acknowledging the concept of changing stakeholder requirements. Both concepts demonstrate that sustainability is not just a simple measure of time but is relative, as it is highly dependent on how **S**, **F** and **T** are defined in a specific context.

Consensus on what sustainability means in the field of software and requirements engineering is still emerging despite a number of attempts to formalise a definition

[43]. However, the Karlskrona Manifesto for Sustainability Design [6] provides a focal point for establishing a common ground for the software and requirements engineering community to engage with sustainability by advocating a set of fundamental principles and commitments that underpin sustainability design. The principles stress the importance of recognising that sustainability is an explicit consideration, even if the primary focus of the system under design is not sustainability. It also advocates that sustainability must be viewed as a construct across five dimensions—environmental, economic, individual, social and technical—and considers the potential long-term effects of systems.

The concept of sustainability has been discussed extensively in a number of publications, and readers are directed to these for an in-depth treatment of this topic [4, 6, 13, 41, 43].

3 Related work

While traditional RE methods and tools do not explicitly facilitate the discussion of sustainability-related concerns, research suggests that existing RE techniques, approaches and methods can serve as a starting point for practitioners to integrate sustainability into their practice [9]. Chitchyan et al. [14] identified several techniques that helped support sustainability in RE and demonstrated the application of some of these techniques using two case studies. Similarly, Mireles et al. [32] proposed a conceptual framework for the classification of sustainability-aware requirements methods to support practitioners in the selection of an appropriate method to address stakeholders' needs. However, the results of their analysis suggested that existing approaches were heavily biased towards sustainability goals related to effects of the ongoing use of systems (rather than structural effects from their long-term use) and the case studies addressed mainly early requirements during the development stage of the software life cycle.

A number of studies have also attempted to integrate sustainability into specific methods and techniques. Seyff et al. [41] extended the WinWin Negotiation Model to consider the effects of requirements on sustainability. The results of the study suggested that while the approach stimulated the discussion across the various dimensions of sustainability, stakeholders found it challenging to identify the effects of a given requirement on sustainability and were not able to identify long-term effects. Cabot et al. [9] proposed using **i*** for modelling early requirements as a way to visualise the impact of alternative options on sustainability goals and to analyse the conflicts between sustainability and other problem-specific objectives. Their approach is based on explicitly representing the sustainability effect of each business or design alternative, in order to allow stakeholders to understand the trade-offs between sustainability and other business goals and making informed decisions. Similarly, Mussbacher and Nuttall [33] argue that goal models are an ideal candidate support the assessment of alternatives for sustainability as they express the hierarchy of needs from high-level goals to specific activities for various stakeholders. However, there has been no comparative evaluation of either Cabot's [9] or Mussbacher's [33] method to demonstrate their efficacy in addressing sustainability.

Brito et al. [7] argue that to properly address the various dimensions of sustainability, approaches need to enable reasoning as well as assess the impact on each other and on other system concerns in the very early stages of software development. To achieve this, they proposed a concern-oriented requirements approach that allows both the modelling of sustainability concepts and their relationships and the management of conflicting situations triggered by impacts among sustainability dimensions or between those and other system concerns. In contrast to the previous studies, Penzenstadler et al. [35] explored how the concept of leverage points could be used to make sustainability issues more tangible in a public transportation system [42], discussion how, within a complex system (such as an economy, a living body, a city or ecosystem), a small shift in one thing can produce big changes elsewhere [31]. The results suggest that while leverage points do not tell us exactly how to act on sustainability challenges, they provide an analysis tool to help practitioners to identify elements that can bring about effective change at different levels, for a (software) system and the wider system it resides in by offering insights on possible transformation mechanisms and/or ways to find alternatives.

A number of other approaches have also been proposed, including the use of a recommender system to overcome the barriers of incorporating sustainability into the software engineering process [37], the application of a sustainability requirement pattern to guide the specification of sustainability requirements [38], a tool for requirement engineers to analyse the impact of the requirements on system sustainability [3] and a meta-model which integrates sustainability dimensions with the other quality attributes [40]. However, none of the above approaches has focused discussion on the role of requirements engineering for sustainability engineering.

4 Sustainability awareness framework (SuSAF)

The SuSAF was developed using Design Science [17], a rigorous process of designing artefacts to solve problems, to evaluate what was designed or what is working, and to communicate the results [10]. The main goal of the framework is to raise awareness of the sustainability effects that a software system could have in its intended context. Awareness of such prospective effects is essential for all stakeholders engaged in system design: from clients who are commissioning the system to IT product managers and others who will be affected by the system implementation [15, 27, 28]. The SuSAF includes a set of instructions, forms and questions that can be used to guide discussions with the stakeholders, either by means of semi-structured interviews with stakeholders or workshops. It is intended to be used by requirements engineers to engage the broad range of relevant stakeholders into discussions on sustainability. These discussions can lead to further analysis by system designers and other stakeholders and, consequently, to changes in the requirements of the system to try to mitigate potential negative effects and exploit positive ones.[1] In the following sections, we provide an overview of how the different elements of the framework were designed. While the SuSAF has evolved through a number of iterations, in this paper, we focus on the first and the last iterations of their development, to which we refer as the "baseline" and "spring 2019". These are detailed in Sect. 5.4.

4.1 Design of question sets

The question sets are the core of the framework. To elicit them, we used an adaptation of the Delphi method [25, 34]. Here, the members of the Karlskrona Alliance on Sustainability Design [5] acted as the panel of experts, as they have worked on topics of sustainability for over six years, focusing on various domains, such as energy, food security and smart cities. Our aim was not to create an exhaustive list of questions to address every aspect of sustainability (which is quite impossible), neither it was to cover all domains and types of systems. Instead, we wanted to provide a practical **starting point** for stakeholders to discuss possible sustainability effects of technical systems. As a result, we deliberately converged on a set of questions that would cover only five topics for each dimension.

To start the process, the facilitator (first author of this paper) set out an online document and invited panel members to contribute views on the main topics to consider for each of the five dimensions of sustainability and questions that the stakeholders should consider regarding these topics. Two example software systems—Airbnb and a procurement system—were used to ground the discussions. Airbnb was chosen as it is a generally well known and commonly used system, whereas the procurement system was studied by the panel of experts in a previously reported work [4]. The panel then worked through three rounds of activities to converge on the final set of topics and questions.

The first (contribution) round started with the panel members providing their views by directly editing a document and populating topics and questions. In this round, the panel members were asked to write down their own contributions, without any other concern. The facilitator closed this round when all the contributors felt they had listed the most important issues. She then reviewed all topics and questions, removed repetitions and rephrased the questions for better readability. She also consulted selected literature (previously suggested by the panel) to refine the questions. These materials then constituted the result of the first round.

At the second (review) round, the panel was requested to review and comment on the results of the first round. Two weeks were allocated for this round, enabling panel members to contribute their views asynchronously. This resulted in a number of issues raised with regard to previously expressed views/proposed questions (e.g. noting unclear statements, pointing out further implications of the noted event/question, restating leading questions, disagreements with the questions, etc.). The facilitator closed this round when all panel members stated that they had completed their reviews.

The third (consensus) round in the question elicitation process started by the panellists reflecting on the feedback given by others, and reviewing their views in this light. The process continued with the clarifications and resolution of the issues raised. This round was carried out through online small group meetings, where two to four panellists met to discuss the raised concerns. The round terminated when all raised issues were resolved, and all panel members were satisfied with the derived set of topics and questions.

The topics selected through this process are listed in Table 1. However, this does not mean that additional (system and domain specific) topics cannot be considered for each dimension as the discussions proceed. We consider the question sets a foundation for discussion and to be an evolving artefact to which experts from multiple domains are invited to contribute to, while users can tailor the existing question sets to fit their needs.

4.2 Instructions and forms for interviews

In the reported study, the questions were used to guide semi-structured interviews with stakehoders. The instructions for the interviewer detail the interview process starting with forms for the consent to record and collect data as well as guiding questions sheets that help stakeholders to consider various sustainability dimensions and orders of effects. The questions sheets for each sustainability dimension contain questions in plain text, examples, reminders and checkboxes to help the interviewer.

[1] As an awareness framework, the detailed analysis of potential effects is currently outside of the scope of the framework.

Table 1 Topics covered by questions in each dimension

Social	(1) Sense of community; (2) Trust; (3) Inclusiveness and diversity; (4) Equity; (5) Participation and communication
Individual	(1) Health; (2) Lifelong learning; (3) Privacy; (4) Safety; (5) Agency
Environmental	(1) Material and resources; (2) Soil, Atmospheric and water pollution; (3) Energy; (4) Biodiversity and land use; (5) Logistics and transportation
Economic	(1) Value; (2) Customer relationship management (CRM); (3) Supply chain; (4) Governance and processes; (5) Innovation and R&D
Technical	(1) Maintainability; (2) Usability; (3) Extensibility and adaptability; (4) Security; (5) Scalability

Table 2 Questions for the social dimension

Social dimension. Interviewer copy. Tick question as you advance in the interview.

Specific questions	Remind participants to consider...
Sense of community	
[ ] Normally people belong to an **organisation**, to an **area** or to a **group of like-minded people**. Can the system affect a person's **sense of belonging** to these groups?	[ ] ... the **user community** and the **local community**. *Say for example; you mentioned an effect on the sense of community of the user. What about the people in the local community?*
Trust	
[ ] Can the system change the trust[a] between the **users** and the **businesses that own the system**? [ ] What about the trust between the **users themselves**?	[ ] ... **user groups** and **other groups in the society**. *Say for example; you mentioned an effect on how people trust the business. What about how other groups in the society that don't interact with the system trust each other?*
Inclusiveness and diversity	
[ ] Can the system affect **how people perceive others**? [ ] Does the system include users with different **background, age groups, education levels**, or other differences? [ ] Does the system cater for these differences? How?	[ ] ... **user groups** and **other groups in the society**. *Say, for example: you mentioned an effect on the perception of the user. What about other groups in the society?*
Equity	
[ ] Can the system make people to **be treated differently** from each other? For example, because the system carries out data analytics or influences human decisions.	[ ] ... **equality of opportunity**[b] and **of outcome**. *Say, for example: you mention how the system gives the same treatment to people, what about taking actions to ensure the outcome for each person can be the same? For example, putting in place support, communicating in different ways, giving access to resources, respecting decisions, recognising, valuing and respecting differences.* [ ] ... **user groups** or **other groups in the society**. *Say, for example: you mentioned how users are treated by the system. Does the system makes other groups in the society to be treated differently or equally?*
Participation and communication	
[ ] Can the system change the way people **participate** in an **organisation** or other **social groups**? [ ] Does it affect the way people **communicate verbally** and **non-verbally**? [ ] Does it affect the way people **create networks**? [ ] Does it affect the way people **form bounds**? [ ] Does it affect the **effort people put in a group work**?[a] [ ] Does it affect the **actions** people take to achieve the goals, projects and tasks of a group? [ ] Does it affect the way people **engage with others**? [ ] Does it affect the way people **support, consider, critique** or **argue** with others?	[ ] ... the user, the **beneficiaries** and other people **affected** by the system. *Say, for example: you mentioned how users change their way to participate or communicate in groups.*

[a]Trust is a "particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before monitor such action [...] and in a context in which it affects his own action [21]

[b]Social loafing

Turn sheet

The sheet also suggests prompts to encourage the interviewee to think further and examples to clarify some of the questions. For example, a prompt could be: "You mention how the system gives the same treatment to people, what about taking actions to ensure the outcome for each person is comparable?". A clarifying example would be: "Systems sometimes enable the co-creation or co-destruction of value when a customer interacts with the business. E.g. [...] when a customer cannot self-serve as expected, her experience is affected [...]. Does the system enable this kind of co-creation or co-destruction of value?" Table 2 exemplifies the interviewer questions sheet for the social dimension. All question sheets can be found in [2].

The instructions also include a set of questions for the interviewee, to help respondents to follow the interview process.

4.3 Extreme scenarios and chains of effects

The questions (exemplified in Table 2) are intended to help uncover possible immediate-, short- and longer-term effects. In order to encourage identification of such effects, the framework complements questions with a simple note-taking form (shown in Table 3) which explicitly draws the attention of the interviewer to documenting the effects and potential chains-of-effects; that is, when effects are linked by causuality, with one effect providing the ground for the other to appear. It should be noted that separate effects are easier to collect in the note-taking phase and further analysis can help combine the effects into chains of effects.

To foster the interviewee to consider long-term effects, the framework suggests posing an imaginary "extreme" scenario, where the intended software system is accepted and used by millions of people worldwide for a long period of time. The interviewee is then invited to reflect on the effect that such a wide-spread, long-term use of the system may have. For example, in the social dimension, we ask: "Imagine that many people worldwide are using this system for decades. Think about how one thing may lead to another, i.e. a chain of effects. E.g. if people feel closer to their neighbours, they may choose to buy from local shops or choose proximity products, which can then foment local businesses, and finally better distribute wealth". Table 4 shows the interviewee's question sheet for the environmental dimension. Note that it does not include the interviewer's prompt, having instead the questions and prompts for asking about an example of an extreme scenario and chains of effects.

4.4 Sustainability Awareness Diagram (SusAD)

The Sustainability Awareness Diagram (SusAD) is a visualisation tool used to highlight the chains of effects. It can serve as a compilation or a discussion facilitation tool. Requirements engineers could, for example, use it to discuss the main concerns of stakeholders (e.g. extracted from the interviews) with the system designers. They could also adopt it to facilitate the discussion during a workshop, as stakeholders go through the set of SusAF questions, by capturing potential effects in the chart and asking stakeholders to reflect on how one effect may lead to another over time and across dimensions, which is likely to lead to the identification of more potential effects.

The SusAD takes the form of an adapted radar chart (Fig. 1) divided into five equal parts, one for each sustainability dimension, and three concentric pentagons that represent the order of effects. The later denotes how the effects can play out over time. From the centre outwards these effects are: (1) immediate, i.e. a direct function of the system or and direct effect of its development, (2) enabling, i.e. arising from the use of a system, or (3) structural, i.e. referring to persistent changes that can be observed at the macro-level [23].

Let us exemplify the use of the diagram using the example of Airbnb as illustrated in Fig. 1. Airbnb offers a peer-to-peer short-term accommodation booking platform [1]. Airbnb allows property owners to rent out their homes or rooms *as an enabling technical effect*. As a result of persistent rental via Airbnb, homeowners earn 55% more than the median long-term renting, which is an *enabling effect* upon individuals. Increased median long-term rent due to reduced long-term rental accommodation stock is a *structural economic effect* of this platform. Finally, the gentrification of primarily non-white localities and increased race separation is its' *structural social effect*. A SusAD would typically have multiple chain-of-effects.

Table 3 Extract of the note-taking form, with sample notes

Topic	Key points—social dimension
Sense of community	Rent room → personal contact → start of friendship → Better sense of community
	Rating system → welcome and helpful
	High use → change house dynamics → children affected
	High use → door codes → less personal contact
	Structural changes to properties
	High use → long-term renters forced out

Table 4 Interviewee's questions sheet for the environmental dimension, with prompts for asking about an extreme scenario

Environmental dimension (Interviewee copy)	
Specific questions	Final questions
Material and resources [ ]Think about the equipment that are part of the system. Which materials may be consumed to **produce** the system? [ ]What about the **use** of the system? For example, supplies. [ ]Does the system change the way people **consume materials**? For example, encourage people to buy more?	Extreme scenario Imagine that many people worldwide are using this or similar system for many years or decades. Think about how one thing may lead to another. For example, if the system encourages people to buy more clothes, companies will produce more, generating more jobs in the developing world, but also creating greater environmental damage. [ ] Looking at this list of key points you mentioned during the interview, can you think of a chain of effects for some of these key points in the extreme scenario above?
Soil, atmospheric and water pollution [ ] Think again about the **equipments** and **supplies** that are part of the system. Does **producing** them generate **waste** or **emissions**? [ ] Does the system itself produces **waste** or **emissions**? Does the system influence how much **waste** or **emissions** people or institutions generate? [ ] Or, alternatively, does it promote (or impair) **recycling**?	
Biodiversity and land use [ ] Can the system affect the **plants** or **animals around it**? [ ] What about **elsewhere**? [ ] Can the system change the **size, use, of composition of the soil around it**? For example, by **occupying** land or by converting land into **cropland**? [ ] What about **elsewhere**?	
Energy [ ] Does the system affect the **production of energy**? [ ] What about the use of energy? For example, it enables or encourages less energy consumption or consumption from renewable sources? [ ] Does the energy to run the system hardware comes from renewable energy sources?	
Logistics and transport [ ] Does the system affects the **need for movement** of **people** or **goods**? [ ] Does the system affect the **means** by which people or goods move? [ ] Does the system affect the **distance** that people or goods move?	

The framework includes a set of instructions to draw the SusAD. Attempting to capture all chains of effects from interviews or a workshop in a single diagram would make it unreadable and be of limited use. Instead, we suggest using it to capture portions of the potential effects that one wishes to discuss with other stakeholders or system designers. Hence, the instructions also suggest variations of the diagram, which can be chosen according to personal preference. Discussing the alternatives is outside the scope of this paper.

5 Study design

This section describes how we have designed the evaluation of the Sustainability Awareness Framework. Our overall goal with this research is to create a framework for designers of software systems to raise the awareness of the potential effects of such systems on sustainability. This particular work represents a first step towards this goal by creating a question-based framework, which was tested first with students. Therefore, the emphasis of this first evaluation is on the clarity and utility of the question sets for eliciting potential effects of software systems on sustainability. For this, the following research questions (RQs) were addressed:

RQ1 Is the framework applicable to systems that are not directly focused on sustainability?[2]

[2] When first introduced the framework, practitioners often ask whether the framework can be applied to a system whose purpose is not related to sustainability. Hence, we decided to add this RQ.

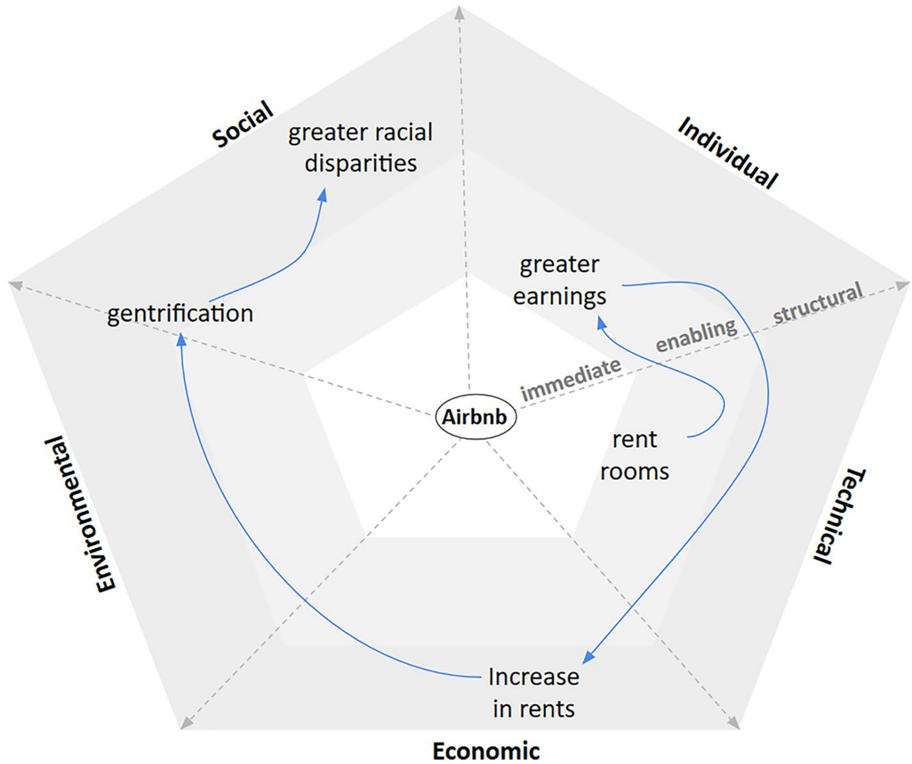

Fig. 1 Simplified SusAD for Airbnb

RQ2 Does the framework encourage insightful discussions about the potential effects of software systems on sustainability?
RQ3 Does the framework help to identify the potential chains-of-effects of software systems on sustainability?
RQ4 Do the questions and the process help to identify **more** effects than just the diagram?
RQ5 Do the questions and the process help to identify **more** chains-of-effects than just the diagram?
RQ6 Is the proposed approach practical?

- RQ6.1 Is it easy to get access to the relevant stakeholders?
- RQ6.2 Are the questions easy to understand?
- RQ6.3 Are the process and the material provided easy to use?

RQ7 How useful is the resulting Sustainability Analysis Diagram (SusAD)?

5.1 Participants

The study consisted of four groups of students using two versions of the Sustainability Awareness Framework, as detailed in Sect. 5.2. Two groups of students participated in the spring of 2018 and two other groups in the spring of 2019. In both years, the studies took place at the California State University, Long Beach (CSULB) and the Lappeenranta University of Technology (LUT). The main difference was that spring 2018 LUT students were studying for a master's degree in the domain of sustainability and therefore did not need an introduction lecture to the topic. Nevertheless, like the other groups, they were not familiar with the SusAF framework.

5.2 Tasks and process

In the spring 2018, the framework was simply the Sustainability Awareness Diagram (SusAD). For this reason, we consider the students on these cohorts as the **baseline**. In the spring 2019, the framework had evolved, being composed of the SusAD, a set of questions about the sustainability dimensions, and a set of instructions and forms for carrying out interviews with the stakeholders and to draw the SusAD.

We next explain the tasks and process for each year, which are summarised in Table 5.

Spring 2018 students: Baseline students were introduced the basic concept of sustainability (including the dimensions and the order of effects) and the SusAD. Then, they were instructed to identify the effects and the chains of effects and to fill out a SusAD for a system of their choice.

More specifically, at CSULB, 26 baseline students worked in 13 groups, in a writing-intensive third-year course on ICT for Sustainability. The assignment was introduced in class after a lecture on the Karlskrona Manifesto [4] and explained with a number of example diagrams.

Table 5 Evaluation study settings

Groups/characteristics	CSULB 2018 (baseline)	LUT 2018 (baseline)	CSULB 2019	LUT 2019
# of students/groups	26/13	21/21	30/9	26/8
Level(BA, MA)	BA	MA	BA	MA
Knowledge of sustainability	Introduction lecture	Degree related to sustainability	Introduction lecture	Introduction lecture
Materials provided	Assignment sheets, lecture slides, background reading	Assignment sheets, background reading [4]	Assignment sheets, lecture slides, background reading	Assignment sheets, lecture slides, background reading [4]
Task set	Identify effects and chains of effects, draw the SusAD	Identify effects and chains of effects, draw the SusAD	Ask questions to interviewees, take notes, identify effects and chains of effects, draw the SusAD only with the most interesting chains of effects	Ask questions to interviewees, take notes, identify effects and chains of effects, draw the SusAD only with the most interesting chains of effects
Measures collected	SusAD	SusAD	SusAD, Questionnaire	SusAD, Questionnaire

In LUT, 21 baseline students worked individually to produce SusADs for their master's thesis. All these students had a good background on ICT as well as sustainability as they studied in the Erasmus Mundus master's programme on pervasive computing and communications for sustainable development. The assignment (based on [4]) was introduced to them while they were preparing their master's thesis).

In both settings, the students had two weeks to create a summary SusAD and to write a report discussing their reflections. This work was carried out by the students independently, and they could approach the researchers in case of questions.

Spring 2019 students: In the following year, students were also instructed to identify the effects and the chains of effects and to fill out a SusAD for a system of their choice. However, they had do it based on interviews with at least one stakeholder per sustainability dimension[3]. For such, they have received a set of questions about the sustainability dimensions, instructions and forms for carrying out interviews and instructions to draw the SusAD. The latter specifically told them to only represent in the diagram the chain-of-effects they found more interesting.

In particular, at CSULB, 31 students worked in nine groups, also in a writing-intensive third-year course on ICT for Sustainability. Of these, 30 students were studying computer science and one environmental management and engineering. The assignment was introduced in class and explained with a number of example diagrams as well as a mock interview with a research assistant.

At LUT, 16 students worked in nine 1–2 person teams. The task was carried out as part of an MSc level course on Sustainability and IT. Here, almost all students (14) had an IT background. The remaining students represented business administration and Sustainability science. The task was introduced in class and explained with several examples on the topic.

In both settings, the students had two weeks to carry out the interviews, create a summary SusAD and write a report, aggregating the information from their interviews and discussing their reflections. This work was carried out by the students independently, and they could approach the researchers in case of questions.

5.3 Data collection

In addition to the data produced by the students from both years, in the spring 2019, we also collected data from the evaluation study by means of two surveys.

The first survey (referred to as "Survey (A)" in Table 8) was answered after each interview with a system stakeholder and contained the feedback of the interviewer and the interviewee about the questions. To collect the impression of the interviewee, students were instructed to ask two additional questions at the end of the interview and to summarise their answers in the online survey. Since the student groups had to carry out at least one interview per dimension, we expected to receive 85 responses. However, we only received 57 (22 from CSULB and 35 from LUT), as some groups did not return all interviews.

The second survey (referred to as "Survey (B)" in Table 8), was answered after each student group filled out the SusAD; it was meant to gather the collective feedback of the group regarding the SusAD framework. It was expected that each group would submit a single response, but some students preferred to submit individual responses. In total, 26 responses were received: 18 from CSULB and 8 from LUT.

[3] Therefore, they interviewed at least five stakeholders and we generally expected that each dimension would be represented by a different stakeholder.

Table 6 Mapping between data sources and RQs

Source	Research questions
Analysis of the SusADs	RQ1, RQ4 and RQ5
Survey	RQ2, RQ3, RQ6 and RQ7

All surveys were collected via Google forms. The diagrams and reports were collected via Dropbox at CSULB and Moodle (a learning management system) at LUT. The survey data were anonymous but reported on the system that the group had been working on. Since the groups were working on a different system, we were able to relate the survey responses to the group work submission.

5.4 Data analysis

The data used to answer the RQs came from both analysis of the SusADs and the surveys, as shown in Table 6. RQ1, RQ4 and RQ5 were answered by analysing the SusADs, while the answers to questions RQ2, RQ3, RQ6 and RQ7 were taken from the surveys.

In the following sections, we explain how the SusADs data were analysed, including an explanation of the analysis of the quantitative and qualitative survey data, and show the relationship between these data sources and the research questions.

5.4.1 Sustainability analysis diagram (SusAD) analysis

The analysis of the Sustainability Analysis Diagram (SusAD) was used to answer RQ1, RQ4 and RQ5. RQ1, which is concerned with the applicability of the framework to different domains/types, is answered by analysing the effects and chains of effects identified for systems of distinct nature. The remaining two questions, which asked whether the question sets helped students to identify **more** effects (RQ4) and chains of effects (RQ5), respectively, were answered by comparing the SusADs from the spring 2018 students (baseline) with the ones from the following year.

In order to analyse the SusADs of both rounds (baseline and spring 2019), the first author of this paper created a codebook with the 26 metrics shown in Table 7. Then, five researchers, including the lecturers and all authors of this paper, extracted the relevant information from the SusADs. Finally, these metrics were mapped to a simple Oppose (O), Inconclusive (I), Support (S) scale to answer RQ1, RQ2 and RQ4.[4]

[4] While these metrics can give us an indication of whether the students could use the framework, a mapping to a five-point scale (as used to answer the remaining RQs) would be unreliable. Hence, we chose a simpler scale.

5.4.2 Survey analysis

We used the surveys to answer RQ2, RQ3, RQ6 and RQ7. In order to do so, we grouped the survey questions in two categories: the first one, shown in Table 8, is composed of questions that directly contribute to answering the RQs, most of which also asked students to provide free text explanations of their choices. The second group contained qualifying questions that helped us to interpret the answers to the first category. Due to space constraints, these are not listed but are discussed in Sect. 6. Finally, for clearer traceability, each survey question was mapped only to the RQs that it most contributed to. We next describe how we interpreted quantitative and qualitative data to answer the RQs.

To analyse the surveys, we used three types of closed questions in the surveys: binary (yes/no), tertiary (yes/no/partially) and a five-point Likert scale. In order to calculate the quantitative results, we mapped responses to numerical values, calculating the median and normalising on a scale from 1 to 5 as follows:[5]

- Tertiary responses (yes/no/partially) were normalised to three-point range in the range 1–5 , where *No* maps to 1, *Partially* maps to 3 and *Yes* maps to 5.
- The five-point Likert scale questions had two cases: some were asked with a "positive phrasing" (e.g. "Have you understood the questions?"), meaning that a higher value would support RQ. These were simply mapped from 1 to 5. Others were asked with a "negative phrasing" (e.g. "Has the interviewee had difficulties in answering the questions?"), meaning that a higher value would oppose the RQ. To calculate the contribution of a negative question, we have used the six complements (i.e. a "2" would become a "4"). In order to differentiate them, "negatively phrased", the type of this questions are marked with brackets, for example "(Likert)".

In order to define the contribution of survey questions to the research questions, we mapped their median values to a scale of support to the questions: Strongly Oppose (SO), Oppose (O), Inconclusive (I), Support (S), and Strongly Support (SS) (Table 9).

Since nearly every survey question also asked to provide a free text explanation for the choice of the scale value, a substantial amount of text was collected for qualitative analysis. This text was coded using a set of codes defined for each survey question, following the

[5] The binary responses (b.4 and b.7 in Table 8) were not used for RQ analysis directly, as they are only qualifying questions for answers to b.5 and b.8, respectively.

Table 7 Information extracted from SusAD

ID	Data	ID	Data	ID	Data
1	Domain	2	Type of system	3	Relation to sustainability
4	Total number of effects	5	Number of positive effects	6	Number of negative effects
7	Number of first-order effects,	8	Number of second-order effects,	9	Number of third-order effects
10	Number of effects in social dimension	11	Number of effects in individual dimension	12	Number of effects in environmental dimension
13	Number of effects in economic dimension	14	Number of effects in technical dimension	15	Median number of effects per dimension
16	Number of empty cells (dimension vs order of effects)	17	Total number of chains of effects	18	Minimum length of chain
19	Maximum length of chain	20	Median length of chain	21	Number of chains that cross dimensions
22	Minimum number of dimensions crossed	23	Maximum number of dimensions crossed	24	Median number of dimensions crossed
25	Number of chains that cross order of effects	26	Median number of chains of effects crossed		

Table 8 Survey questions that directly contributed to RQs

Survey (a): After each interview—interviewer		Type
a.1	Have you understood the questions?	Likert
a.2	Has the interviewee had difficulties in answering the questions?	(Likert)
a.3	Did the questions enable discussions with the interviewees?	Likert
a.4	Did you get insightful answers using the questions in this particular domain?	Likert
a.5	Have the interviewees been able to think of chains of effects for the extreme scenario?	Likert
Survey (b): After the SusAD—group		**Type**
b.1	Did you get access to the right stakeholders? [1 per dimension]	y/no/p
b.2	How easy was it to extract possible effects from the discussion prompted by the questions? The conversation might have been unstructured or messy; thus, how easy was it to extract the information needed from it?	Likert
b.4	Did the questions help to fill out the SusAD?	Likert
b.4	Did you use the form for taking notes on the key points of the interview?	y/n
b.5	If yes, how do you like the form for taking notes on the key points of the interview?	Likert
b.6	How did you like the process of asking the questions, noting down the key points and showing them back to the interviewee?	Likert
b.7	Did you refer to the Drawing Instructions for drawing the Sustainability Analysis Diagram (SusAD)?	y/n
b.8	If you read them, where they useful?	Likert
b.9	Was the resulting SusAD readable?	Likert
b.10	Was the resulting SusAD useful?	Likert
Survey (a): Questions to Interviewee		
a.6	Were the questions easy to understand?	Text
a.7	Have the questions been useful for triggering relevant discussions on the possible effects of software systems? Why?	Text

Table 9 Mapping of the values to the scale of support

Scale	Value range
Strongly oppose (SO)	1
Oppose (O)	2
Inconclusive (I)	3
Support (S)	4
Strongly support (SS)	5

qualitative content analysis approach [29]. Two researchers created the codebook and collaboratively analysed the free-text responses from CSULB, after which the coding was validated by two additional researchers. The codebook was reused for analysis of survey data from LUT. Double coding was used if a given free-text response related to several code categories. For example, the answer “I would need to interview more people with different expertise” would be coded as “more_interviews” and “different_expertise”. Table 10 shows an extract

Table 10 Examples of codes for survey questions

RQ	SQ	Cont.	Codes	Occur. A	Occur. B
RQ1	a.7	pos.	Useful_to_expand_on_topic	10	12
		neg.	Useful_some_not_relevant_for_system	2	3
		pos.	Useful_for_future	2	5
		pos.	Useful_because_detailed	2	2
		neg.	Useful_insufficient_expertise	1	2
		pos.	Useful_because_vague	1	1
		pos.	Useful_to_take_action	0	1
		pos.	Useful	0	4
		pos.	Useful_privacy	1	0

from the codebook, which contains the RQ it refers to, the survey question, whether the contribution of the code towards the question is "positive" or "negative", the code, and the number of occurrences for both universities. These qualitative data were used for interpreting the respondents' choices.

5.4.3 Mapping data sources to research questions

Table 11 shows how each data source was used to answer each RQ. Note that by having two surveys, one conducted after the interview and another conducted after depicting the SusAD, allows us to have three data points for drawing conclusions on some of the questions: the feedback of interviewees, interviewers and the group as a whole.

5.5 Threats to validity

Threats to validity hamper the ability to draw conclusions from the evidence [45]. For the Delphi study, the main risks come from the fact that the members of the panel were drawn from the Karlskrona Alliance on Sustainability design, who are experts in software and requirements engineering, not in individual sustainability dimensions. Yet, they have been working on sustainability-related topics for several years. Thus, the breadth of the views represented in the question sets is biased towards the group's own view and may not coincide with the views of experts on the sustainability dimensions. Furthermore, the anonymity of the panel members was not preserved (as they know each other). This could cause a number of additional biases, where attitudes towards the individuals could have influenced the agreement or disagreement with their provided views. In the future, this might be mitigated by integrating input from experts from outside the group and validating the question set through wider participation.

For the evaluation study, one of the main risks is the reactive bias, as the students might answer the questionnaire positively to meet the expectations of their teachers (i.e. halo effect). Additionally, there are several confounding factors which may affect the outcome that were not taken into account, such as differences in knowledge regarding sustainability issues of the students and the level of expertise of the interviewees. Since we worked with four different groups of students from two different universities, these factors cannot be ruled out completely. However, we endeavour to ensure a similar perspective on sustainability and knowledge of the questions and the SusAF by delivering introductory sessions on sustainability and instructions to both groups. Another risk is concerning the comparison of the baseline with the spring 2019 students. In the first round, we did not

Table 11 Mapping between RQs and survey questions

ID	Research question	Source
RQ1	Is the framework applicable to systems that are not directly focused on sustainability?	Analysis of SusADs
RQ2	Does the framework encourage insightful discussions about the potential effects of software systems on sustainability?	Survey questions: a.3, a.4, a.7
RQ3	Does the framework help to identify potential chain-of-effects of software systems on sustainability?	Survey questions: a.5
RQ4	Do the question and the process help to identify **more** effects than just the diagram?	Analysis of SusADs and comparison with a past course (the "baseline")
RQ5	Do the questions and the process help to identify **more** chains-of-effects than just the diagram?	Analysis of SusADs and comparison with a past course (the "baseline")
RQ6	Is the proposed approach practical?	See below
RQ6.1	Is it easy to get access to the relevant stakeholders?	Survey questions: b.1
RQ6.2	Are the questions easy to understand?	Survey questions: a1, a.2, a.6
RQ6.3	Are the process and the material provided easy to use?	Survey questions: b.2, b.3, b.4, b.6, b.8
RQ7	How useful is the resulting Sustainability Analysis Diagram (SusAD)?	Survey questions: b9, b.10

recommend the students to only picture the most interesting effects in the SusAD. Thus, the comparison is biased through this recommendation.

Another main risk is the possible bias caused by result interpretation. We applied researcher triangulation and mixed qualitative and quantitative methods to minimise this risk.

Finally, we do not attempt to generalise the findings from these two application cases; we only demonstrate the feasibility of using the SusAD for relating requirements engineering process to topics of sustainability.

6 Results

This section presents the results of our data analysis and explains how the results address the RQs. Table 12 summarises the extent to which the analysis of the different SusADs and surveys provided evidence for answering our RQs. Average quantitative results that led us to the final conclusions about the support to a given concept are shown in brackets.

6.1 RQ1: framework applicability

Research question RQ1 investigated whether the framework was applicable to systems that are not directly focused on sustainability. To answer this question, we review the *breadth of the domains* of the systems to which the framework has been applied and then detail the differences between the systems which are *directly related to sustainability* and those that are not. We examine this question as it is a common concern when analysts and students are first introduced to the framework.

We consider a system to be *directly related to sustainability* if the *main purpose* of the system is to foster some sustainability goal(s). Examples of systems that are directly aimed at fostering sustainability are a system enforcing sustainable agricultural practices and a system enabling control of renewable energy generation.

Other systems, such as Amazon Kindle and Uber, are not directly aimed at a specific sustainability goal, but have a *clear and immediately discernible sustainability effects*, e.g. Amazon Kindle removes the need for cutting trees for book printing and Uber allows car owners to earn additional income and passengers to get cheaper, more transparent and faster transportation services. We consider such systems to be *arguably related* to sustainability.

Finally, other systems, such as Netflix and Wolf—a food order and delivery system—do not have either explicit or immediately discernible sustainability goals. We consider such systems to be *not related to sustainability*. Note, this is only to say that there is no immediately visible link from such systems to any sustainability goal. It is not to say that these systems are irrelevant for sustainability or that sustainability is irrelevant to them.

In terms of the **breadth of domains**, students were free to choose any system of interest to them. Analysis of the SusAD produced showed that the students were able to use the framework for systems of different types and domains. With respect to the type of system, nearly half (55.5%) were pure Information Systems (IS) and 37.5% were both Embedded and Information Systems (ES & IS). The remaining two were a Cyber-Physical System (CPS) and a non-software system. The domain varied widely, with the most common ones being transportation (27.8%), entertainment (22.2%) and food (16.7%). The remaining domains—household products, energy, virtual reality, education, games and health care—only had one instance each. Table 13 shows the systems that the students have chosen and their respective classification. From this, we can clearly see that the framework has been applied to a wide variety of domains and system types. Thus, neither the domain nor the system type has acted as barriers for the use of the SusAF.

With regard to their **relationship to sustainability**, when considering the direct and immediately discernible sustainability goals of the considered systems, we note that 27.8% of systems were clearly related to sustainability, 27.8% could be arguably related to sustainability, and 44.4% were not related at all (as shown in the last column of Table 13).

For all types of systems, students were able to find potential effects. It is curious to note that they found even more effects in systems that were classed as unrelated to sustainability, as shown in Table 14.

We also observed that the more the system's purpose is related to sustainability, the higher the percentage of positive effects identified. This might be attributed to the fact that these systems have been designed to improve the effects on the different dimensions of sustainability. However, interestingly, even for systems that have not been built with sustainability as their key objective, the percentage of positive effects is still greater than the negative ones (as shown in Table 14).

When it comes to the distribution across the order of effects, we observed no particular difference between the three types of system. Neither could we observed anything significant with respect to the distribution of effects between the dimensions or the median number of effects per dimensions. Effects were similarly distributed in all three cases.

The relation to sustainability also did not affect the average number of chains of effects that were found. However, we do observe that the systems more related to sustainability had somewhat longer chains of effects; these differences are not very substantial. We have also observed that the average number of chains of effects that crossed dimensions is larger for systems whose purpose have some

Table 12 Summary of the results

RQ	Sub-RQ	University	Source				
RQ1			Analysis of SusADs				
		CSULB	S				
		LUT	S				
		Supported					
RQ2			Individual Survey				
			a.3	a.4	a.7		
		CSULB	S	S	I		
		LUT	S	S	S		
		Supported					
RQ3			Individual Surveys				
			a5				
		CSULB	I				
		LUT	S				
		Supported					
RQ4			Comparison of SusADs				
		CSULB	S				
		LUT	S				
		Supported					
RQ5			Comparison of SusADs				
		CSULB	O				
		LUT	O				
		Oppose					
RQ6	RQ6.1		Group Survey				
			b1				
			Soc.	Ind.	Eco.	Env.	Tec.
		CSULB	SS	SS	I	I	SS
		LUT	I	I	I	O	S
		I					
	RQ6.2		Individual Surveys				
			a1	a2	a6		
		CSULB	S	S	SS		
		LUT	S	S	SS		
		S					
	RQ6.3		Group Surveys				
			b.2	b.5	b.6	b.8	b.3
		CSULB	S	S	S	SS	S
		LUT	I	I	O	S	S
		S					
	Supported						
RQ7			Individual Surveys				
			b.9	b.10			
		CSULB	S	S			
		LUT	S	I			
		Supported					

Table 13 Classification of the systems domain, type—Cyber-Physical System (CPS), Embedded System (ES), Information System (IS) and others—and relation to sustainability

System	Domain	Type	Sustainability purpose
California State University, Long Beach (CSULB)			
Reusable water bottles	Household products	No software	Yes
Solar energy	Energy	CPS	Yes
Amazon Kindle	Entertainment	Hybrid (ES & IS)	Arguably
Electric Scooters	Transportation	Hybrid (ES & IS)	Arguably
Self-driving cars	Transportation	Hybrid (ES & IS)	Arguably
Gacha games	Entertainment	IS	No
Virtual reality	Virtual reality	Hybrid (ES & IS)	Arguably
Hyperloop	Transportation	Hybrid (ES & IS)	Arguably
Sustainable agriculture	Food	Hybrid (ES & IS)	Yes
Lappeenranta University of Technology (LUT)			
Netflix	Entertainment	IS	No
Elder scrolls online, add-on	Games	IS	No
Uber	Transportation	IS	No
YouTube	Entertainment	IS	No
ResQ Club	Food	IS	Yes
E-prescription (medicine)	Health care	IS	No
Duolingo	Education	IS	No
Wolt	Food	IS	No
BlaBla Car	Transportation	IS	Yes

Table 14 Metrics with respect to relation to sustainability

Relation to sustainability	Average number of effects	Percentage of positive effects (%)	Percentage of negative effects (%)	Median length of chain	Average of chains crossing order of effects	Average of chains crossing order of effects
Related	19.5	76.9	23.1	3.5	6.0	9.3
Arguably related	16.0	62.5	37.5	7.0	7.2	3.5
Unrelated	21.4	58.7	41.3	4.4	9.0	3.5

relation to sustainability (35% and 58% for related and arguable related), but not the number of chains that crossed the order of effects. These metrics are summarised in Table 14.

The analysis of the SusADs suggests that students successfully managed to apply the framework to systems of different domains and types. While we observed some differences related to the direct relatedness of the system's goal to sustainability, these differences are not significant. We therefore conclude that the evidence **supports RQ1**.

6.2 RQ2: insightful discussion

Our second research question (RQ2) was interested in whether the framework encouraged insightful discussions about the potential effects of software systems on sustainability?

Three survey questions contributed to this RQ; the answers we received are summarised below. The first survey question asked whether the SusAD questions **enabled discussions with the interviewee** (a.3) and the students from both CSULB and LUT **supported** (4) this notion. Analysing the qualitative codes for that question (shown with occurrences in parenthesis), we observed that some students stated that the questions led to more questions (17%, "lead_to_more_questions" CSULB=7 LUT=3) and helped to elaborate the answers (10%, "elaboration" CSLUB=2 LUT=4). For example, one respondent said: *"[...] the interviewee could direct the direction of topic and voice his personal opinions without influence from us"*. Furthermore, the questions were reported to be supportive (9%, "good_support_from_questions" CSLUB=7 LUT=3), and to encourage an interviewee who is knowledgeable (12% "knowledgeable_interviewee" CSLUB=2 LUT=3) or passionate (5% "enable_passionate_interviewee" CSLUB=2 LUT=1); e.g. "*This topic is something the interviewee was very passionate about*". We note a difference in the textual answers received from two universities, which could be due to the cultural difference in communication in California vs Finland: while in CSULB only one interviewee was described as terse, in the LUT, six interviewees and one interviewer received such description. This could also be observed in the difference in the number of codes generated for the data from these two universities. (Due to space constraints, we will no longer show the codes and occurrences related to qualitative findings, but all explanations to students' choices have been analysed in the same way as above.)

The second survey question asked whether **insightful answers** (a.4) had emerged using the questions in this particular domain. Again, the students **supported** (4) this notion, with support from CSULB and LUT. The most cited reason for getting good insights was the interviewee opening new perspectives about the domain (14%), followed by having a lot to discuss (7%): *"The questions explored areas I would not have thought of on my own."* and *"The applicable questions were very insightful and invoked lots of back and forth discussion"*. The most frequent reason for not getting much insight was insufficient domain knowledge (7%). Students got the best insights into the individual dimension, followed by environmental and technical. No dimension was particularly problematic. To get more information on insights, we also asked students whether **anything unexpected** came up. Only 26% reported unexpected occurrences, the most common being new perspectives (8%) and the effects of the system (8%).

Finally, the third survey question asked the interviewees perceptions on whether the questions had been useful for **triggering relevant discussions** on the possible effects of the software system (a.7). Interviewees **supported** (4) this notion, with CSULB being inconclusive and LUT supporting it. For example, one student mentioned, *"Yes, we discussed many topics triggered by the questions asked"*. A majority of students confirmed that the interviewees found the questions helpful to expand on the topic, to think towards the future and to discuss in more detail. Helpful pointers towards exploring environmental aspects and privacy were mentioned. The reasons for reporting less usefulness were that some questions were not relevant for that particular system (9%) and that the interviewees did not consider themselves sufficiently knowledgeable (5%).

The answers suggest that these two studies **support (4) RQ2**. That is, that the questions enabled relevant discussions, both for the interviewers and interviewees, and led to insightful findings.

6.3 RQ3: identifying chains-of-effects

Research question RQ3 investigated whether the framework helped to identify the potential chains-of-effects of software systems on sustainability. When asked whether interviewees had been able to **think of chains of effects** for the extreme scenario (a.5), the overall answer was **supported** (with a supported by LUT students but inconclusive for CSULB). To explore this further, students were asked for how many topics (in Table 1) interviewers were able to think of chains-of-effect: 78% thought of chains of effect for up to three key topics, 8.5% for four–five topics and only 3.5% for more than five topics.

Table 15 Correlation of encouragement and ease to think of chains of effect (smaller number = greater difficulty & less encouragement)

Ability to think of chains-of-effects	Encouraged to think about the extreme scenario	Encouraged cross dimensional thinking
1–2	2.39	1.78
4–5	4.03	3.19

Table 16 Correlation of encouragement and number of topics with chains-of-effect (smaller number = less encouragement)

Normalised number of topics	Encouraged to think about the extreme scenario	Encouraged cross dimensional thinking
1	2.8	2.1
2	3.4	2.7
3	4.2	3.7
4	5.0	4.4

To see whether the extreme scenario helped interviewees to think of chains-of-effect, we asked whether the students had encouraged the interviewee to think about the extreme scenario and about effects across dimensions. We observe that those with difficulties in identifying chains-of-effect were less encouraged to think about the extreme scenario and cross-dimensional effects. The opposite is also true, as shown in Table 15. In addition, the more the interviewee was encouraged to think of an extreme scenario and across dimensions, the greater the number of topics (s)he identified chains-of-effect for. This correlation is shown in Table 16. These suggest that the extreme scenario and the encouragement given by the interviewee are indeed useful to identify chains-of-effect.

Finally, we also asked for how many topics (in Table 1) interviewers were able to think of chains-of-effect. 78% of the interviewees thought of chains of effect for up to three key topics, 8.5% for four–five topics and only 3.5% for more than five topics. Interestingly, the more the interviewee was encouraged to think of an extreme scenario and across dimensions, the greater the number of topics he or she identified chains-of-effect to. This was also observed by students, who stated that "*Giving them to consider of chains of effects allows for their thought process to expand past just one dimension*" and "*All these things are interrelated and are* *necessary to examine when researching a topic like this*". Finally, around 30% of the students admitted to not having asked the interviewee to consider the extreme scenario. The primary reasons for not doing so varied greatly. The most cited reasons were the difficulty of including the questions (5%) and that the extreme scenario was not relevant for the system (5%).

The answers suggest that these two studies **support RQ3**. That is, the questions help to identify effects and chains-of-effects, highlighting the importance of the extreme scenario and the encouragement to think across dimensions.

6.4 RQ4: identify more effects

Research question RQ4 investigated whether the questions and the process help to identify more effects than just the diagram. To answer this question, we compared the spring 2019 students' SusADs with SusADs produced in a previous course, when the question sets and the process did not exist yet (the baseline). This accounted for 15 additional group SusADs for CSULB and 21 for LUT.

With the question sheets, we observed an increase in the number of effects identified in both universities, with a 10% increase in CSULB diagrams and an 80% increase in LUT diagrams. While it is impossible to verify the correctness of the effects identified by interviewees—this is a reflection exercise on non-existing systems—this indicates that the questions broaden the students' perspectives, helping them to identify more potential effects. This is particularly evident for the LUT students, who had more previous knowledge on sustainability, but were previously unable to clearly relate that knowledge to the software systems development.

Also, while baseline CSULB students often reused a few of the effects they had seen in the example SusADs used to explain the method, spring 2019 students came up with more unique SusADs that were more specific to their respective system.

It is worth emphasising that while we gave no instructions to the baseline students with respect to the number of effects that they should represent in the diagram, in the spring 2019 round the instructions specifically recommended students to draw only the most interesting effects. Hence, the overall increase in the number of effects could have been even greater, if not for this recommendation. However, further research is needed to investigate this issue in more detail.

We observe that, in both universities, baselines students tended to find more positive effects than negative. Only 17% of effects found by baseline CSULB students were negative, while this number for baseline LUT students was as little as 2%. We interpreted that students tended to recognise effects that reinforce the benefits of the system. With the questions, however, spring 2019 students became more critical, classifying in both universities 35% of effects as negative.

When the distribution of effects between first, second and third orders are compared, we note that in the baseline cohort, the students had these effects similarly distributed. With the questions (spring 2019) there was some increase in the number of second-order effects (+16% for CSULB and +25% for LUT) and a similar decrease in third-order effects (-13% for CSULB and -19% for LUT), while the number of first-order effects remained roughly the same. This may indicate that the framework questions are implicitly related to the second order effects, or that an increase in the number of second-order effects reduced the time available for further third-order effects explorations. We note that the study of the third-order effects is directly supported through the extreme scenario tool of SuSAF. Thus, the analysts using SuSAF may refocus the tool use (allowing more time on detailed discussion of questions or extreme scenario) based on the timeline they wish to focus on (mid- vs longer-term future).

On the other hand, we did not observe any significant difference in the distribution of effects across sustainability dimensions or in the number of effects per dimension. There was also no observed difference between the average numbers of empty cells (i.e. effects that are not considered for one dimension in a particular order of effect). This suggests that, with or without the questions, the students followed their own perceived concerns across time.

As such, the data suggest that when compared to the baseline diagrams, the students who used the SuSAF question sets and application process were able to think of more effects, and so their perspectives on sustainability had broadened. Also, the baseline students often simply reused the example effects given by the lecturers or carried out a limited exploration of the effects that reinforced the systems' purposes. In contrast, spring 2019 students from both universities developed both a larger number of effects and were more uniquely related to their studied systems. Therefore, we conclude that the evidence extracted from the SusADs **supports RQ4**.

6.5 RQ5: more chains-of-effects

Research question RQ5 investigated whether the questions and the process helped to identify more chains-of-effects than just the diagram. We answered this RQ by comparing the current SusADs with the SusADs produced by the baseline students.

Interestingly, in both universities, baseline students found twice as many chains of effects and the medium length of these chains was longer (2 times for CSULB and 1.5 for LUT). We also observed that in the baseline diagrams, chains crossed twice as many dimensions and order of effects as with the questions.

We note several possible reasons for this. In the first instance, as noted before, discussing the questions themselves had drawn on attention and substantial time for the spring 2019 students, leaving little opportunity for thinking about the chains of effects. Specifically, the extreme scenario (which helps to think about the chains of effects) was discussed only towards the very end of the interviews. A substantial number of students did not budget enough time to discuss the scenario with the interviewees at all. This clearly reduced the input into the chain identification process.

Secondly, the 2019 students were explicitly instructed to draw only the most interesting chains-of-effects, while the baseline students received no such instruction. As a result, the representations of findings from 2019 students could have been rather selective and limited.

Finally, the reason could be attributed to the design of the instrument itself, and a better process of framework application may be needed to help focus not only on the effects but also on their chains.

Further research is needed to narrow down the cause of the reduced chain length and to encourage more long-term thinking.

Evidence shows that the baseline students found more and longer chains of effects, which crossed more dimensions and orders of effects. Therefore, the evidence **opposes RQ5**.

6.6 RQ6: framework practicality

The primary aim of research question RQ6 was a preliminary check if the framework is suitable for testing with industry. It looks into access to expertise, the ease of the process and the usefulness of the materials. As a result, we divided this research question in three sub-questions, which are discussed in the following.

6.6.1 RQ6.1: stakeholder access

When we asked whether the students felt that they had been able to **access to the right stakeholders**, the overall answer was **inconclusive**. There were “strong support” from CSULB and a “mixed” response from LUT. Of the five dimensions, both groups found it easiest to get access to relevant stakeholders for the individual and technical dimensions. Experts for the economic dimension proved the hardest to obtain for both groups.

As a result, the answer to this question was **inconclusive**, with students from CSULB finding it easier to access the right stakeholders than the LUT ones.

6.6.2 RQ6.2: ease of understanding

The first survey question asks whether the **interviewee understood the questions** (a.1). Responses from both universities **supported** (4) this idea. However, about one-fourth of the students pointed out that there were questions with unclear definitions, which points out the need to review and refine some of the questions. The questions that caused greatest confusion were related to the supply chain (23%) and agency (11%). No dimension was particularly problematic. However, 10% of the students felt that some questions were not relevant to the system at hand. This was expected; in creating a general framework, we knew we could neither be specific to a domain nor comprehensive.

The second question in the survey refers to the **interviewer’s perceptions** on whether **interviewee could understand the questions** (a.2). The answer to this question was a **support** (4).

The greater difficulties reported by students were that some questions had no relation with the system (15%); e.g. one student said: *“Difficult to get a conversation going about the topic, the interviewee did not consider there to be much to discuss”*. Other reported difficulties were lack of knowledge of the interviewee (10%) and the wording of the questions (9%). To get a deeper understanding of their answers, we asked whether students had interviewed an expert or a surrogate. Overall, about 70% of the interviewees were surrogates, and 30% had knowledge or expertise on the topic. We found little correlation between the level of expertise and the observed difficulty in understanding the questions.

The third survey question refers to **interviewee’s view** on whether the questions were **easy to understand** (a.6). Interviewers from both universities **strongly supported** (5) this notion. One interviewee mentioned, *“Yes, the questions got me thinking about change and the decisions we have to make for a sustainable future”*. Furthermore, responses indicate that some questions were perceived as vague (12%), again showing the need for reviewing and refining some questions. Other interviewees felt that some questions were not relevant for the system under analysis (9%), or they needed time to be interpreted (7%).

The answers to these questions suggest that both interviewers and interviewees understood the questions. Therefore, the overall these answers **support RQ6.2 (4)**.

6.6.3 RQ6.3: ease of use

Five survey questions contributed to this RQ. The first one asked how easy it was to extract possible effects from the discussion prompted by the questions, considering that the conversation might have been unstructured or messy (b.2). Overall, students' answers were a support (4), with CSULB students' supporting (4) this notion and LUT's answers being inconclusive (3). Explanations varied quite widely and for a broad range of reasons. Around 10% of the students considered that the questions were helpful to extract the desired effects. For example, one student said: *"Questions were straight to the point and made it easy to get their opinions"*. Furthermore, 14% had positive feedback on how the conversation with the interviewee had helped in this regard. Limiting factors were hard questions (7%), the limited amount of public information (2%), hard to organise information (2%) and that the interviewee did not answer some of the questions.

The second question asked whether students liked the note-taking form. (b.5) However, to qualify the answers to this question, we checked how many of them have actually used it. While a slight majority (62%) was using the form provided by us for taking notes, a significant minority of 38% did not. CSULB students were more diligent in using the form (66%) than LUT ones (50%). The main reasons the subjects gave for not using the form was that they did not find it useful (12%), they had forgotten to bring it (8%) or they had made their own sheet (8%). One student, for example, said that s/he preferred *"working on one same sheet rather than switching between the question sheet and the answer sheet"*. Overall, students who used the form supported (4) it, with data from CSULB showing support (4) and from LUT being inconclusive (3). Some students felt the form was helpful in several ways (37.5%), while others felt it had limited space (6.2%) and enforced working with multiple sheets rather than one (6.2%).

The third survey question refers to whether the students **liked the process** of asking the questions (b.6), noting down the key points and showing them back to the interviewee. Data were **inconclusive** (3) about this research question, with students from CSULB supporting it (4) and from LUT opposing (2) this notion. The reasons why subjects liked and disliked the process varied greatly. On the positive side, the subjects liked interviewing (12%) and found it to be a good practice (8%). The main reason for disliking it was that subjects felt it was sometimes redundant (12%).

The fourth question asked whether the students thought the instructions for drawing the SusAD were useful (b.8). The instructions were liked by the majority of the students (87%), 2 did not answer and 2 disliked it. The overall result was that the groups supported (4) this notion, with CSULB strongly supporting (5) it and LUT just supporting it (4). Respondents highlighted that the instructions were clear (25%) and helpful to guide their drawings (50%). For example, one student stated that the *"instructions were really clear with definition and example of every step"*. Some students commented on the clarity/usefulness of the examples (5) and others on the fact that the instructions were particularly helpful to those drawing SusADs for the first time (4). A couple of students made more generic comments, simply stating they were useful and helped to remind them of what had been learned in class. The two who criticised the instructions felt that they were long and challenging to understand.

The last survey question asks whether the **questions helped to fill out** the Sustainability Awareness Diagram (b.3). Respondents from both universities **supported** (4) this notion. Some students found the questions helpful for the diagram (9%) and to extract key points (7%): *"Answers were straight forward, so the points were easy to establish on the SusAD"*. Furthermore, some saw that questions allowed them to extract key points (7%) and helped the interviewee to generated further ideas (4%). In addition, they saw the questions as good way to start (1%), were straightforward (2%) and showed continuity (2%). Only two respondents mentioned a negative effect in that certain questions could lead to bias (2%) and that they were too direct (2%).

Finally, considering the length of the interviews, we note that most of them took between 15 and 30 min (52%), some others lasted from 30 to 60 min (27%) and a few took less than 15 min (12%). Even though the interviewees who identified most chains-of-effect had also talked for longest, no clear correlation was found between the time of the interview and the number of chains-off-effect identified. It could well be that the shorter interviews were given by more knowledgeable and less available stakeholders.

The answers to these sub-questions RQ6.3 suggest that they support (3.76) RQ6.3. While we could not conclude that it was easy to extract chains of effects from the discussion prompted by the questions, students felt that the questions helped to fill out the SusAD. Furthermore, they generally liked the note-taking form, the instruction to draw the SusADs, and the proposed interview process, though several opportunities for improvement were also noted.

The overall answers to these questions **support (4) RQ6**; that is, students found the framework practical. Although we could not conclude whether they were able to access the right stakeholders, students found the questions easy to understand and both the process and material easy to use.

6.7 RQ7: sustainability analysis diagram (SusAD) usefulness

The first survey question asks whether the resulting **SusAD was readable** (b.9). Students from both universities **supported** this idea (S). The three main explanations were that the diagram was readable (23%), the students decided

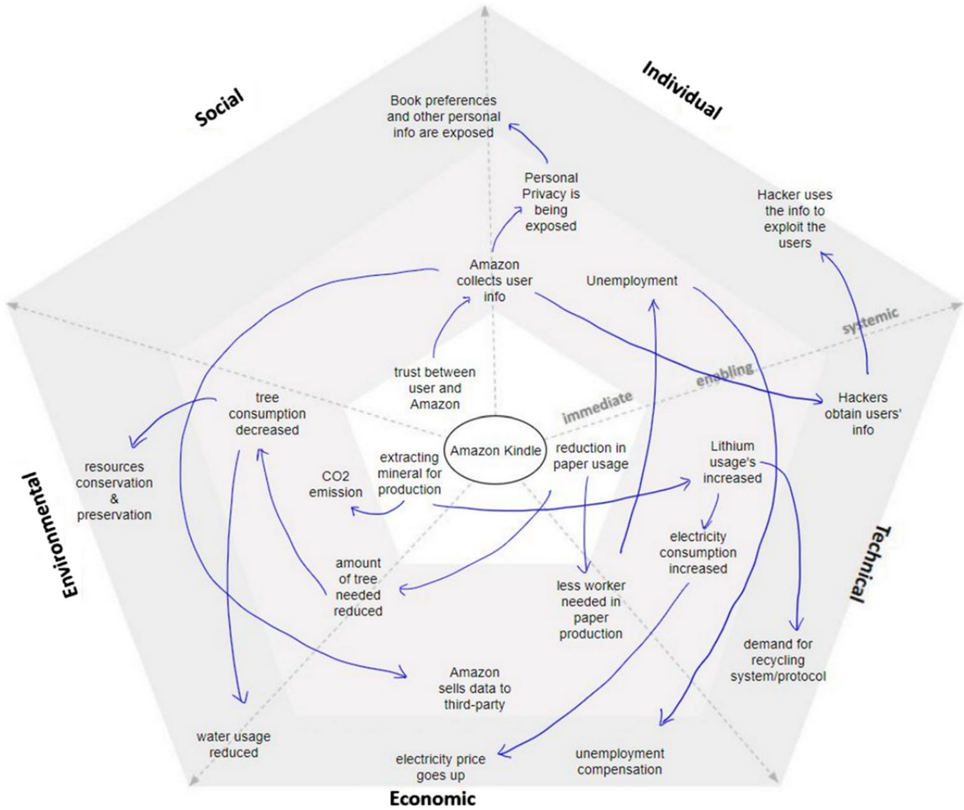

Fig. 2 SusAD produced by CSULB students on Amazon Kindle

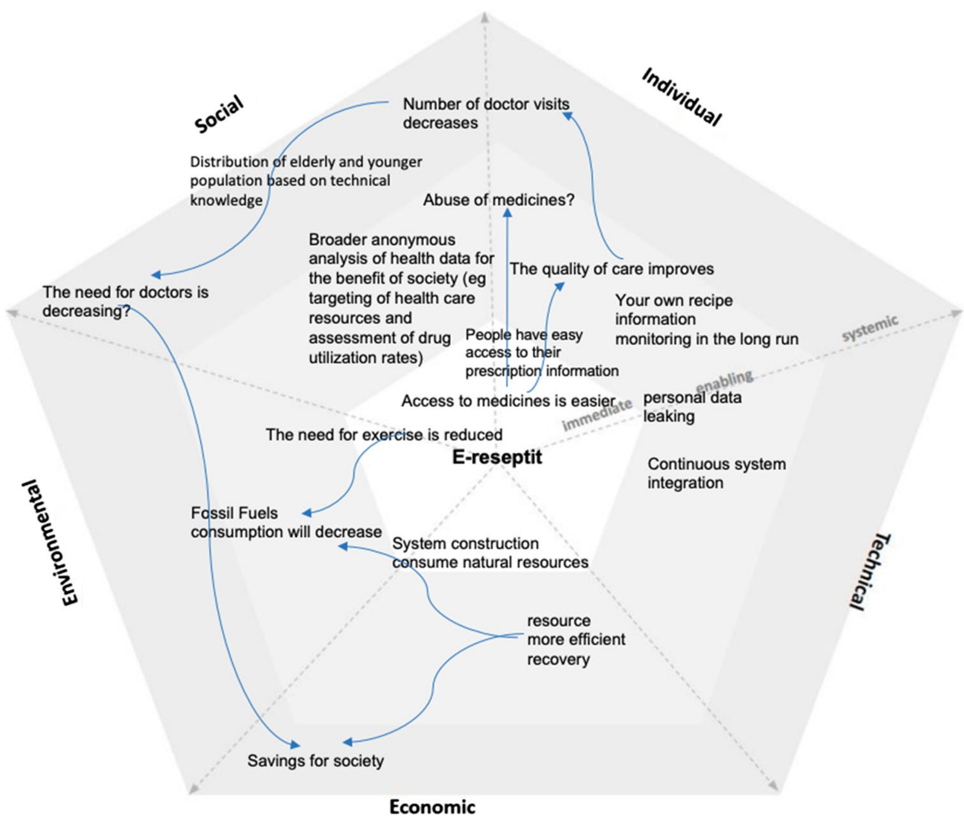

Fig. 3 SusAD produced by LUT students on a e-prescription system

only to include key points (15%) and that they were able to make links (12%). For example, one student said that *"once we made it, it was easy to understand our effects"*, while another one stated, *"we made sure to avoid cluttering the diagram with unnecessary information"*. Figures 2 and 3 exemplify SusADs produced by students at CSULB and LUT, respectively.

When asked whether the resulting SusAD was **useful** (b.10), overall students **supported** that notion (4). While the answer from LUT students was inconclusive (3), CSULB students felt that the SusAD was helpful for writing their essays for an accompanying assignment (33%), helped them visualise (11%) effects and was easy to understand (11%). From example, one student said *"It organised all our thoughts visually in a way that is easy to understand"*.

The answers suggest that these two studies **support (4) RQ6**. That is, students felt that the SusADs were readable and useful.

Overall, the evaluation study supports six out of seven research questions. The framework is applicable to different types of systems and domains; it encourages insightful discussions about the potential effects of software systems on sustainability, it helps to identify effects (especially when the extreme scenario is used), is practical, and the resulting SusAD is useful. Yet, there is still room for improvement, in particular in the questions for the social and economic dimensions, and in the interview process.

7 Discussion

In the following sections, we discuss the lessons learned based on instructors' reflections on the process and procedures of applying the framework, the analysis, as well as our general observations, which are discussed in the following sections.

7.1 Instruction

For the instructors participating in this study, it was generally easy for them to apply the framework as they were part of the group who developed the approach (see also researcher bias in threats to validity). Therefore, it is difficult to judge the ease by which an independent facilitator could successfully apply the framework. However, as a partial indicator, one of the instructors had only been partially involved in the development of the framework and did not report any difficulties with instructing the students.

7.2 Method

Sufficient time must be set aside for describing the approach as it is not self-explanatory, which is a common point for instruction in general. When the students really know what is expected and how to achieve that, their analysis becomes much better. We found it helpful to use classroom roleplay where an instructor and a teaching or research assistant would play the interviewer and the domain expert. Roleplay is a common method in teaching approaches that involve stakeholders [20, 46, 47], and very beneficial in this case, as the students could see how to use the framework. Having used the framework with and without question sets, we observed that using the questions helped the students to understand the dimensions and the orders of effects as well as making them feel more confident when conducting the sustainability effect analysis, without worrying about whether they had the sufficient expertise.

7.3 Stakeholder challenge

A big challenge the students was to find adequate surrogate stakeholders. Students had much richer set of results when talking to either domain or product experts. For example, for the economic dimension, they would approach students studying for their Masters degree in Economics. In other instances, students used their friends (who had relevant experience with the chosen systems) as surrogate stakeholders. Where the students had someone, with a relevant background, to help think through the sustainability concerns with respect to their chosen, their work resulted in good analyses.

7.4 Wording

Some of the students paraphrased the questions into their own words. They explained to us that they were concerned that the questions, as formulated in the question sets, could be hard to understand. When analysing the results with and without paraphrased questions, we cannot see any noticeable difference. Thus, what matters is that the students and their interviewees understood the questions and could answer them to identify possible sustainability effects. However, it is clear that there is scope to simplify further the SusAF questions for the future improvement of the framework.

7.5 SusAF as a systems thinking activity

The proposed Sustainability Awareness framework incorporates key concepts from the field of Systems Thinking [30] into the RE process. Our work advocates consideration of the holistic system within which the software-to-be will function, attending not only to the functional and non-functional properties of the software system but also to the indirect, longer-term effects that its use could cause including emergent behaviour, and the risks and uncertainties that this may engender. However, we are also aware that the discipline of software engineering already suffers from high costs and late delivery problems [44], and additional "whole systems" analysis could prove too costly and complex to be useful. In truth, this very problem stifles the use of techniques such as Soft Systems Methodology [12] or Critical Systems Thinking [22] in the software engineering domain. To avoid unbounded complexity and cost, our approach supports the exploration of potential sustainability effects through a guiding set of question and an effect recording tool for an elicitation scenario. This allows the focus to remain on the identification of sustainability effects across the three orders of effect and provides a boundary to the potentially overwhelming systems thinking and analysis task.

7.6 Systems versus software requirements engineering

It has long been recognised that requirements engineering is one of the most critical parts of the software and systems development life cycle, which can determine the success or failure of a product or project [26]. Requirements engineers working within the systems engineering domains (such as construction and chemical process engineering) are well attuned to conducting systemic impact analysis. However, this is too often amiss within the field of software engineering, where RE has too often limited itself to the elicitation of software requirements, disregarding the wider implications that the software could cause [6, 41]. It is our hope that this paper has sufficiently articulated (using the Airbnb platform example) the need to tackle such disregard of the socio-technical effects of software systems. In this paper, we propose a framework to take the first steps in addressing this omission.

7.7 Requirements engineers as leads for sustainability engineering

Researchers have previously argued [4] that requirements shape the software systems, which in turn shape the socio-technical systems within which they reside. As such, if so engineered, software systems could become the drivers towards sustainable societal, environmental, and economic settings. As a result, the present work endeavours to support requirements engineers in taking on the role of sustainability engineers, through timely consideration and fostering informed choices in tackling the challenge of the socio-technical systems requirements engineering.

8 Conclusions

This paper highlights that the software developed today does not exist in isolation but forms part of the wider socio-technical system within which it gets used. This paper advocates that requirements engineers must tackle concerns of sustainability of such socio-technical systems, during the software requirements elicitation phases of a software system life cycle. In conclusion, we found evidence that the Sustainability Awareness Framework presented in this paper provides an accessible approach to elicit effects that software systems could have. That is, it could be used by students independently and without previous knowledge. As previous research argues that students are a close enough representation for practitioners [24, 39], we are hopeful that the framework will serve equally well for professionals.

Having evaluated the questions that guide such awareness-building activity with two sets of student groups, we find sufficient evidence that the questions and elicitation scenario provide the desired support. However, much work remains to be done. Importantly, we need to provide better support for the identification of potential chains-of-effects. We also plan to improve the clarity of some of the questions, to consider the specialisation of the questions per relevant application domains, to develop digital tools for supporting the use of the framework (e.g. note-taking, analysis and the visualisation of chain-of-effects, all of which are currently done manually), to support the evaluation of alternatives and to study the use of the framework across different cultures.

In addition, while we have rigorously collected data for evaluating the SuSAF, we have not followed any set evaluation framework. However, a framework such as TAM [16] would allow our results to be better compared against another established technique. Similar approaches could also include the application of the goal-question-metric (GQM) approach, which could be helpful in structuring the goals of the study and its evaluation process. Finally, to overcome the limitation of the Delphi method explained in Sect. 5.5, we will continue to invite additional, external experts on the different dimensions to comment on and expand the question sets.

Acknowledgements Each named author made a significant contribution in terms of ideas, discussions, evolution and physical writing to be named on the paper. All authors participated in the design and development of the framework applied in the study. The authors would also like to thank Dr. Christoph Becker for insights that led us to change the focus from an analytic to an awareness framework and Ms. Nanae Aubry for helping with part of the coding of the feedback forms.

References

1. Airbnb, inc. http://www.airbnb.com 2019. Accessed 01 March 2019
2. Akinli Kocak S, Becker C, Betz S, Chitchyan R, Duboc L, Easterbrook S, Leifler O, Penzenstadler B, Porras J, Seyff N, Venters CC (2019) The SusA Workshop—improving sustainability awareness to inform future business process and systems design. https://doi.org/10.5281/zenodo.3632486

3. Alharthi AD, Spichkova M, Hamilton M (2018) Susoftpro: sustainability profiling for software. In: IEEE 26th international requirements engineering conference (RE), pp 500–501
4. Becker C, Betz S, Chitchyan R, Duboc L, Easterbrook S, Penzenstadler B, Seyff N, Venters C (2016) Requirements: the key to sustainability. IEEE Softw. 33(1):56–65
5. Becker C, Chitchyan R, Duboc L, Easterbrook S, Mahaux M, Penzenstadler B, Rodriguez-Navas G, Salinesi C, Seyff N, Venters C, Calero C, Akinli Kocak S, Betz S (2015) Karlskrona manifesto on sustainability design. https://www.sustainabilitydesign.org/
6. Becker C, Chitchyan R, Duboc L, Easterbrook S, Penzenstadler B, Seyff N, Venters C (2015) Sustainability design and software: the karlskrona manifesto. In: Proceedings of the 37th international conference on software engineering-volume 2, IEEE Press, pp 467–476
7. Brito IS, Conejero JM, Moreira A, Araújo J (2018) A concern-oriented sustainability approach. In: 2018 12th international conference on research challenges in information science (RCIS), pp 1–12
8. Brundtland GH, Khalid M, Agnelli S et al (1987) Our common future. New York
9. Cabot J, Easterbrook S, Horkoff J, Lessard L, Liaskos S, Mazon J (2009) Integrating sustainability in decision-making processes: a modelling strategy. In: 31st international conference on software engineering—companion volume, pp 207–210
10. Cagdas V, Stubkjær E (2011) Design research for cadastral systems. Comput Environ Urban Syst 35(01):77–87
11. Checkland P (1999) Systems thinking. Rethinking management information systems, pp 45–56
12. Checkland P, Poulter J (2006) Learning for action: a short definitive account of soft systems methodology and its use, for practitioners, teachers and students. Wiley, Hoboken
13. Chitchyan R, Becker C, Betz S, Duboc L, Penzenstadler B, Seyff N, Venters C (2016) Sustainability design in requirements engineering: state of practice. In: Proceedings of the 38th international conference on software engineering companion, ACM, pp 533–542
14. Chitchyan R, Betz S, Duboc L, Penzenstadler B, Easterbrook S, Ponsard C, Venters C (2015) Evidencing sustainability design through examples. In: Proceedings of the 4th international workshop on requirements engineering for sustainable systems
15. Damian D (2007) Stakeholders in global requirements engineering: lessons learned from practice. IEEE Softw. 24(2):21–27
16. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q:319–340
17. Dresch A, Lacerda DP, Antunes JAV (2014) A method for science and technology advancement. Springer, Berlin
18. Duboc L, Betz S, Penzenstadler B, Akinli Kocak S, Chitchyan R, Leifler O, Porras J, Seyff N, Venters CC (2019) Do we really know what we are building? Raising awareness of potential sustainability effects of software systems in requirements engineering. In: 2019 IEEE 27th international requirements engineering conference (RE), pp 6–16
19. Easterbrook S (2014) From computational thinking to systems thinking: A conceptual toolkit for sustainability computing. In: ICT for Sustainability, (2014) ICT4S-14. Stockholm, Sweden, p 2014
20. Gabrysiak G, Giese H, Seibel A, Neumann S (2010) Teaching requirements engineering with virtual stakeholders without software engineering knowledge. In: 2010 5th international workshop on requirements engineering education and training, IEEE, pp 36–45
21. Gambetta D (2000) Can we trust trust? In: Gambetta D (ed) Trust: making and breaking cooperative relations, vol 13. Oxford University Press, Oxford, pp 213–237
22. Goodchild M, Janelle D (2010) Toward critical spatial thinking in the social sciences and humanities. GeoJournal 75(1):3–13
23. Hilty L, Aebischer B (2015) Ict for sustainability: an emerging research field. In: ICT innovations for sustainability, Springer, pp 3–36
24. Höst M, Regnell B, Wohlin C (2000) Using students as subjects—a comparative study of students and professionals in lead-time impact assessment. Empir Softw Eng 5(3):201–214
25. Landeta J (2006) Current validity of the Delphi method in social sciences. Technol Forecast Soc Change 73(5):467–482
26. Laplante PA (2017) Requirements engineering for software and systems, 3rd edn. Auerbach Publications, Boca Raton
27. Laukkanen E, Paasivaara M, Arvonen T (2015) Stakeholder perceptions of the adoption of continuous integration—a case study. In: 2015 agile conference, pp 11–20
28. Levy M, Hadar I (2018) Multidisciplinary requirements engineering for addressing social-oriented concerns. In: 2018 1st international workshop on learning from other disciplines for requirements engineering (D4RE), pp 1–3
29. Mayring P (2014) Qualitative content analysis: theoretical foundation, basic procedures and software solution
30. Meadows D (2008) Thinking in systems: a primer. Chelsea Green Publishing, Hartford
31. Meadows DH (1999) Leverage points: places to intervene in a system
32. Mireles GAG, Moraga M Á, García F, Piattini M (2017) A classification approach of sustainability aware requirements methods. In: 2017 12th Iberian conference on information systems and technologies (CISTI), pp 1–6
33. Mussbacher G, Nuttall D (Aug 2014) Goal modeling for sustainability: the case of time. In: IEEE 4th international model-driven requirements engineering workshop (MoDRE), pp 7–16
34. Okoli C, Pawlowski S (2004) The Delphi method as a research tool: an example, design considerations and applications. Inf Manag 42(1):15–29
35. Penzenstadler B, Duboc L, Venters CC, Betz S, Seyff N, Wnuk K, Chitchyan R, Easterbrook SM, Becker C (2018) Software engineering for sustainability: find the leverage points!. IEEE Softw 35(4):22–33
36. Penzenstadler B, Raturi A, Richardson D, Tomlinson B (2014) Safety, security, now sustainability: the nonfunctional requirement for the 21st century. IEEE Softw 31(3):40–47
37. Roher K, Richardson D (2013) A proposed recommender system for eliciting software sustainability requirements. In: 2013 2nd international workshop on user evaluations for software engineering researchers (USER), pp 16–19
38. Roher K, Richardson D (2013) Sustainability requirement patterns. In: 2013 3rd international workshop on requirements patterns (RePa), pp 8–11
39. Runeson Per (2003) Using students as experiment subjects–an analysis on graduate and freshmen student data. In: Proceedings of the 7th international conference on empirical assessment in software engineering, Citeseer, pp 95–102
40. Saputri T, Lee S (2016) Incorporating sustainability design in requirements engineering process: a preliminary study. In: Asia pacific requirements engineering conference, Springer, pp 53–67
41. Seyff N, Betz S, Duboc L, Venters C, Becker C, Chitchyan R, Penzenstadler B, Nöbauer M (2018) Tailoring requirements negotiation to sustainability. In: 2018 IEEE 26th international requirements engineering conference (RE), pp 304–314
42. UK Commission for Integrated Transport. Transport and climate change: advice to government from the commission for integrated transport. http://www.cambridgeenergy.com/archive/2007-02-08/commission-integ-trans.pdf (2007)
43. Venters C, Seyff N, Becker C, Betz S, Chitchyan R, Duboc L, Mcintyre D, Penzenstadler B (2017) Characterising sustainability requirements: a new species, red herring, or just an odd fish? In: IEEE/ACM 39th international conference on software engineering, Institute of Electrical and Electronics Engineers Inc., 6, pp 3–12

44. Verner J, Sampson J, Cerpa N (2008) What factors lead to software project failure? In: 2008 second international conference on research challenges in information science, pp 71–80
45. Wohlin C, Runeson P, Hst M, Ohlsson M, Regnell B, Wessln A (2012) Experimentation in software engineering. Springer, Berlin
46. Zowghi D, Paryani S (2003) Teaching requirements engineering through role playing: lessons learnt. In: Proceedings of 11th IEEE international requirements engineering conference, 2003., IEEE, pp 233–241
47. Zuppiroli S, Ciancarini P, Gabbrielli M (2012) A role-playing game for a software engineering lab: developing a product line. In: 2012 IEEE 25th conference on software engineering education and training, IEEE, pp 13–22