User Personas Improve Social Sustainability by Encouraging Software Developers to Deprioritize Antisocial Features

Bimpe Ayoola, Miikka Kuutila, Rina R. Wehbe and Paul Ralph
Faculty of Computer Science, *Dalhousie University*, Halifax, NS, Canada
bimpe.ayoola@dal.ca, miikka.kuutila@dal.ca, rina.wehbe@dal.ca, paulralph@dal.ca

Abstract—Background: Sustainable software development involves creating software in a manner that meets present goals without undermining our ability to meet future goals. In a software engineering context, sustainability has at least four dimensions: ecological, economic, social, and technical. No interventions for improving social sustainability in software engineering have been tested in rigorous lab-based experiments, and little evidence-based guidance is available. *Objective*: The purpose of this study is to evaluate the effectiveness of two interventions—stakeholder maps and persona models—for improving social sustainability through software feature prioritization. *Method*: We conducted a randomized controlled factorial experiment with 79 undergraduate computer science students. Participants were randomly assigned to one of four groups and asked to prioritize a backlog of prosocial, neutral, and antisocial user stories for a shopping mall's digital screen display and facial recognition software. Participants received either persona models, a stakeholder map, both, or neither. We compared the differences in prioritization levels assigned to prosocial and antisocial user stories using Cumulative Link Mixed Model regression. *Results*: Participants who received persona models gave significantly lower priorities to antisocial user stories but no significant difference was evident for prosocial user stories. The effects of the stakeholder map were not significant. The interaction effects were not significant. *Conclusion*: Providing aspiring software professionals with well-crafted persona models causes them to de-prioritize antisocial software features. The impact of persona modelling on sustainable software development therefore warrants further study with more experience professionals. Moreover, the novel methodological strategy of assessing social sustainability behavior through backlog prioritization appears feasible in lab-based settings.

Index Terms—Sustainable development, Social sustainability, software engineering, sustainable software engineering, persona, stakeholder map

I. INTRODUCTION

Sustainable Development refers to meeting the needs of the present without compromising the ability of future generations to meet their own needs [1]. Sustainable software engineering, analogously refers to creating and maintaining software in a way that meets present needs without undermining our collective capacity to meet our future needs [2]. Sustainable software development therefore involves minimizing negative impacts on the economy, society, human beings, and the environment during software development, deployment, and usage [3]. Put another way, sustainable software development means allowing technology to improve our daily lives while mitigating social risks (e.g. cybercrime, surveillance capitalism, social isolation) [4].

Software sustainability encompasses four dimensions: *economic* sustainability concerns wealth creation and profitability; *environmental* (or *ecological*) sustainability relates to the responsible use of natural resources; *technical* sustainability involves designing quality software for adaptability [5]; *social* sustainability relates to impacts on the wellbeing of individuals, organizations, communities, and society [2]. The social dimensions is the most complex and difficult to define of the four.

Sustainability is rarely addressed holistically [6]. Researchers and practitioners often focus on environmental or technical sustainability of software [7]–[13], but rarely address social sustainability despite sometimes being considered as important [2], [6]. This gap exists partly because social sustainability is qualitatively complex and lacks clear criteria, practical guidance, or tools for integrating it into software development processes [2]. For instance, a software development team may address environmental sustainability by enhancing energy efficiency, technical sustainability through refactoring, and economic sustainability through market research and user-centered design (emphasizing a feasible revenue model).

However, integrating social sustainability is more challenging. Development is socially sustainable when it meets present needs while improving, or at least not undermining, human wellbeing. Social sustainability involves creating software that benefits society, intersecting social issues such as equity, social justice, privacy, security, human well-being [14]. It takes different forms at different levels; for example, individual physical and mental health; group cohesion; organizational inclusiveness, community resilience, national social justice, or world peace. Therefore, development may be unsustainable in many ways; for instance, by deleteriously affecting safety, equity, tolerance, democratic participation, human rights, the rule of law, education, or sustainability awareness.

Socially sustainable software development requires not only understanding the societal impact of software products but also embedding principles that support social well-being and ethical responsibilities into the development process. This aspect of sustainability, often overshadowed by its environ-

mental and technical counterparts, demands a more nuanced approach—delving into social values and impacts such as equality, diversity, community building, and belongingness.

The challenges in incorporating sustainability into software development are further highlighted by a recent meta-analysis of sustainable software engineering research [2]. This analysis revealed a critical gap: only a few studies have evaluated sustainability-promoting interventions [11], [15]–[17], and none of these studies used controlled experiments. While controlled experiments are not the *only* way to evaluate an intervention, the absence of any controlled experiments is a red flag. This underscores the need for more research to develop effective strategies for integrating social sustainability in software.

This study begins to address this gap by evaluating the role of stakeholder mapping and personas in contributing towards social sustainability. WE pose the following research question.

> **Research Question:** *To what extent do stakeholder maps and personas affect how software professionals' prioritize social sustainability?*

Stakeholder mapping is the process of identifying and categorizing stakeholders by considering their needs, their concerns, and how a product or project might affect them [18], [19]. Personas, on the other hand, are fictional characters created based on research to represent different user types. They help developers represent intended users to understand their needs, goals, characteristics and diversity [20]–[22]. They can also be used as a tool to facilitate sustainable user behaviour in software design [23].

II. RELATED WORK

A. Social Sustainability

Social sustainability focuses on maintaining social capital and preserving the solidarity of communities [5], as well as ensuring equal or improved access to social resources for future generations, highlighting the importance of generational equity [24], [25]. This concept is critical in the context of software development, where socially sustainable software should aim for digital inclusion by eliminating barriers, promoting multicultural awareness, and making systems more accessible to a diverse user base. This includes improving accessibility for the elderly, people with disabilities, non-English speakers, and low-literacy users to provide equal opportunities to the users [26].

However, accounting for the social dimension of sustainability is challenging for practitioners [27] due to its intangible, qualitative nature and lack of consensus on relevant criteria [28], [29]. There is no clear method or comprehensive set of metrics for evaluating the social sustainability of a software product [30]. Existing guidelines assessing social sustainability are often vague and inadequate [31]. In a recent survey, zero of 178 software professionals reported considering social sustainability in their development process or being guided by an explicit sustainability policy [32].

Significant attention has been directed towards the societal impacts of technology, particularly how AI technologies can affect marginalized and vulnerable populations [33], [34]. The increasing integration of AI in software systems brings unique challenges to social sustainability. An exploratory study [31] on the negative social impacts of existing AI systems highlights concerns such as discrimination, effects on people with disabilities, socio-economic division, and inequality in opportunities. Issues range from discriminatory AI in health predictions to a variety of threats including malicious use, trust erosion in authorities, and gender inequality. The study also emphasizes the lack of user awareness regarding AI discrimination and underscores the need for developing guidelines to promote socially sustainable AI technologies. In addressing these challenges, the role of human-centered design (HCD) practices becomes paramount, especially those that promote social justice. Traditional HCD approaches often center the experiences of dominant groups, inadvertently pushing vulnerable populations behind [35]. There is a need to shift towards more inclusive design practices, and implementing a social justice framework to prevent marginalization and promote inclusive design [35].

Anticipatory measures, which involve proactive efforts to predict and prepare for potential future developments and risks associated with innovations, are also increasingly recognized as essential in technology governance. Stilgoe's framework for Responsible Innovation, which identifies anticipation, inclusion, reflexivity, and responsiveness as its core dimensions, emphasizes the importance of foresight in understanding and addressing the societal implications of emerging technologies [36]. Anticipation, in particular, encourages stakeholders to consider the long-term societal implications of technological advancements. The Responsible Innovation framework further highlights that technology and society are deeply interconnected, and as such, innovation must align with societal values and ethical standards. Embedding anticipatory practices not only addresses immediate concerns but also anticipates future societal impacts, ensuring that technological advancements contribute positively to society [37].

Al Hinai and Chitchyan [4] identified over 600 social sustainability indicators in their systematic literature review, grouping them into 12 categories including including "employment", "health", "education", "human rights", and "equality". Their subsequent work [38] focused on equality, suggesting that recognizing value patterns can help software engineers document the social sustainability requirements of the software under development.

Most recently, Moises de Souza et al. [39] conducted a systematic literature review identifying various social aspects of sustainability in software development. These include "diversity", "trust", "equality", "security", "transparency", "human rights", and "fairness". They highlight the need for software practition-

ers to adopt practices that promote these social goals, such as fostering inclusive environments, upholding ethical standards, protecting human rights, and promoting transparency and fairness. The review also identified tools and practices aimed at fostering awareness of societal implications and ensuring alignment of software development with social sustainability goals, although they noted that these tools require further empirical validation to ensure their practicality and effectiveness in real-world software development scenarios. Tools like the Sustainability Awareness Framework [13] aim to raise software engineers' awareness of the impacts of their work on societal sustainability. However, the effectiveness of these tools and practices still requires further empirical validation. McGuire et al. [2] advocate for more research evaluating the effectiveness of sustainability interventions.

B. Stakeholder Mapping

An organization's *stakeholders* are individuals or groups who can affect or are affected by its objectives or achievements [40]. Analagously, a project's stakeholders are individuals or groups who can affect or are affected by the project and its outcomes. *Stakeholder mapping* is the process of visualizing an organization or project's stakeholders; the visualization is called a *stakeholder map*. Stakeholder maps may also include information about stakeholder's values, interests, relationships to one another, or how the product might affect them [19].

Johann et al. [41] emphasize the pivotal role of user and community engagement in fostering sustainable software engineering. Their work underscores the significance of socio-cultural contexts and promotes including diverse user backgrounds throughout the development process. This perspective resonates with the broader understanding that software products relies on several stakeholders including end-users. These stakeholders are not always fully identified or considered during software development and often times, only the most obvious or prominent ones are taken into account [19], [42].

Several studies have identified stakeholder identification as a tool that promote sustainability [18], [19], [43], [44]. Johann et al. [41] also pointed out the importance of thinking about where users come from and their backgrounds when making software products, ensuring everyone is included. However, while these tools and methods sound great for making software more sustainable, there is still more work to do in validating and assessing them in a practical software development environment.

C. Personas

Personas are fictional characters used to describe user characteristics, goals, attitudes, and experiences when using software products [45]. Personas go beyond mere statistical representations, descriptions, or basic information of average users; they weave together actual user behavior patterns, motivations, and frustrations with fictional information such as names, images,

and backgrounds thereby rendering them more tangible and relatable [46].

Personas enhancing the cognitive processes of designers, and foster a user-centric approach to design [47]. Without the guidance of personas, designers tend to over-rely on their personal experiences or assumptions. Personas help anchor product design in the actual needs of users, mitigating the risk of poorly informed design choices and subjective judgments by software engineers [48].

Creating accurate personas is time-consuming [49], and several methodologies have been proposed [50]–[52]. Some of these studies have also offered insights into effective persona integration in requirement engineering processes [48], and agile methodologies [53].

Despite their benefits, many practitioners still struggle to adopt personas, view personas with skepticism, or perceive them as too abstract, misleading, or irrelevant [54], [55]. This could be due to a lack of familiarity among software practitioners or the absence of validated studies or practical guidelines for using personas effectively [48]. This gap underscores the need for further research evaluating and validating personas with software practitioners to enhance the understanding of user needs in software engineering.

III. METHOD

To address our research question, we carried out a randomized, controlled factorial experiment with 79 undergraduate computer science student participants. After signing consent forms, participants prioritized a backlog of prosocial, antisocial and neutral user stories based on a case description. The process began with a 5-minute instructional video, after which they were asked to begin the prioritization task. One group received a stakeholder map, one group received a set of user personas, one group received both, and the fourth (control) group received neither. All of the materials discussed below are available in our comprehensive replication package (see Section VI).

A. Hypotheses

Our hypothesis are as follows:

H1A: Priorities assigned to prosocial user stories are higher when a stakeholder map is presented.

H1B: Priorities assigned to antisocial user stories are lower when a stakeholder map is presented.

H2A: Priorities assigned to prosocial user stories are higher when a set of persona models is presented.

H2B: Priorities assigned to antisocial user stories are lower when a set of persona models is presented.

We chose stakeholder maps and personas because they are frequently used by software professionals and have been shown to affect decision making [56]–[58]. Stakeholder maps not only help decision-makers identify key actors and those
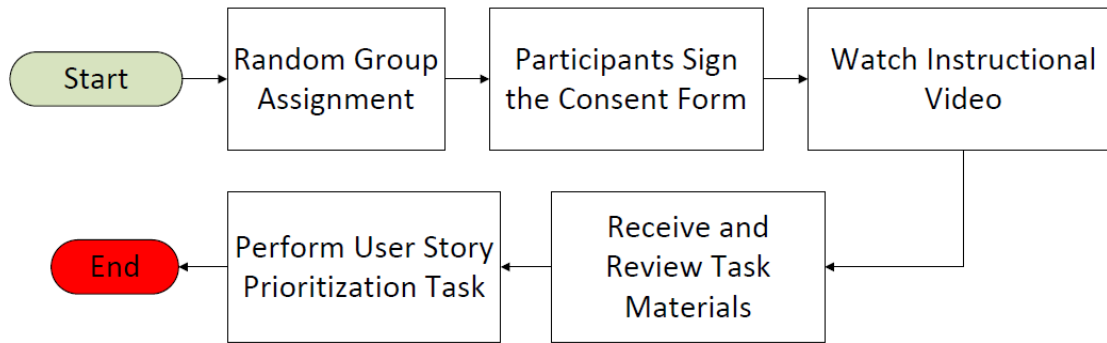
Fig. 1. Flow chart for the experiment process

impacted by a product but also increase the salience of their needs and perspectives [59], [60]. *Salience*—the prominence of information—plays a pivotal role in decision-making as it affects how information is perceived and prioritized [56], [57]. Similarly, persona models deepen understanding of user needs, drive emotional responses, and foster empathy thereby influencing decisions and promoting the consideration of user needs [58]. Persona models and stakeholder maps not only make relevant information more prominent but also act as cognitive enhancers by directing attention to the diverse needs of these groups, thereby improving decision-making outcomes [61].

Moreover, both stakeholder maps and personas can enhancing the visibility of underrepresented groups and implicitly encourage professionals to consider their needs, which is crucial for achieving equitable outcomes. Simultaneously, personas deepen the understanding of end-user contexts, vital for developing features that are socially beneficial. When both personas and stakeholder maps are used in tandem, they could synergistically enhance software professionals' comprehension of user needs. Our factorial experimental design allows us to investigate the interaction effect when both tools are present.

B. Task Development, Instrumentation, and Pilot Testing

We generated a case description involving a facial recognition system (because of the ethical challenges intrinsic to such systems) at a shopping mall (because we think most software professionals and software engineering students in our region are familiar with shopping malls). We then generated a backlog of user stories representing potential features of the facial recognition system.

We conducted a pilot study comprising four rounds of:

- six individuals from diverse cultural background categorized each story as prosocial, neutral, or antisocial;

- for each story with less than 100% agreement, we revised, removed, or replaced the story.

After four rounds we reached 100% agreement on 27 user stories: 9 prosocial (features that would benefit social justice); 9 antisocial (features that would harm social justice), and 9 neutral (with no obvious implications for social justice). The number of stories, 27, was carefully chosen to present a realistic workload without overwhelming the participants or degrading interest. We included neutral stories because a backlog with only pro- and anti-social stories, and no middle ground, would be unrealistic.

In real life, of course, stories could occupy a continuous spectrum or have simultaneous pro- and anti-social effects. The pilot process may have eliminated stories that were borderline or had mixed effects, and intentionally so. Including borderline or mixed-effect stories would decrease objectivity and reliability (by hindering consensus in the pilot process) while potentially making our statistical analysis intractable.

We then implemented the experimental task on paper to encourage more deliberate and thoughtful decision-making without distractions. All participants were given the shopping mall / facial recognition case to provide context. Participants in the experiment groups were given a stakeholder map and/or personas for the prioritization task and instructed to consider the stakeholder map, personas, or both, in assessing the potential impact of each feature on various stakeholders or persona models as they carry out the user story prioritization. Participants in the control group received only the case with neither intervention. Participants were asked to assign each story one of four priorities—1 (don't implement), 2 (low priority), 3 (medium priority), 4 (high priority)—by ticking the appropriate level (see Fig. 2). The list of user stories was also deliberately rearranged and mixed up for each participant to reduce selection bias and also ensure that the order of presentation did not influence their prioritization decisions. This rearrangement involved manually copying and pasting the user stories in different sequences.

C. Recruitment

We used the GPower [62] (version 3.1) to estimate the required sample size for this study. Assuming a medium effect size of

Prioritization Worksheet

Put an X in the block corresponding to the priority you want to assign.

Valid: ☐ ☒ ☐
Invalid: ☐ ☒ ☐

| User Stories	1 – Don't Implement	2 – Low Priority	3 – Medium Priority	4 – High Priority
As a store owner, I want to track shoppers' movements to determine where they go after they leave my shop.				
As a mall manager, I want the facial recognition system to identify persons in distress, to assist mall staff with providing necessary help.				
As a marketing manager, I want to use facial recognition technology to analyze shoppers' emotional responses to advertisements on the digital interface to assess the effectiveness of my marketing strategy.				

Fig. 2. Sample Prioritization Worksheet
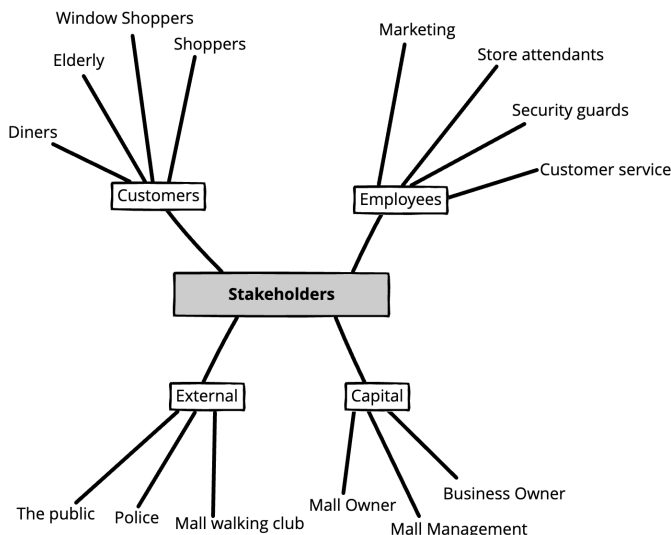
Fig. 3. Sample stakeholder map given to the participants

0.4, 95% confidence level, and an alpha level 0.05 yielded an estimated sample size of 104.

Participants were recruited directly from an undergraduate computer science course at the authors' university. Participating in the study was one of two assignments students could complete for 2% of their term grade. The options were described in the syllabus and participants were recruited through announcements made through the learning management system. The two assignment options were clearly explained, and the alternative assignment was less onerous than participating

in the experiment so as to avoid coercion. Of the 92 students enrolled in the course, 81 attended the experiment session and 79 completed the experiment task (see Table I).

D. Protocol

We used Excel's RAND function to assign each invited participant to one of four groups: control (no stakeholder map or personas), stakeholder map only, personas only, or both stakeholder map and personas—see Table I. Each group was assigned to one of four different rooms on campus, which were selected for their similar layouts and setups to maintain consistency in the experimental environment. We then emailed their room assignments (and reminded them about the right to withdraw and opportunity to complete the alternative assignment).

Each experiment group had two invigilators: the first and second authors and six computer science graduate students. Before the experiment, all invigilators received a detailed briefing from the first author. The briefing focused on the experiment's objectives and procedures, providing step-by-step guidance for the experiment day. Invigilators also received a written experimental protocol.

The experiment was conducted simultaneously with all participants on November 23rd, 2023. All participants were asked to sign a consent form and complete a demographic questionnaire including questions about age, gender, and nationality (see Table II). Participants were then asked to watch an instructional video, which had been pre-recorded by the first author. The video describes the study task and materials. After watching the video, participants were then asked review the

TABLE I
GROUP PARTICIPANT COUNTS

	Invited	No show	Withdrew	Participants
Control	23	3	0	20
Stakeholder map	22	3	0	19
Persona	23	5	2	16
S. map * Persona	24	0	0	24
Total	92	11	2	79

TABLE II
PARTICIPANT SUMMARY

Dimension	Attribute	n	%
Gender	Female	15	18.99
	Male	59	74.68
	Prefer not to say	5	6.33
Age	15-20	22	27.85
	21-26	50	63.29
	27-32	2	2.53
	Prefer not to say	4	5.06
Country	Canada	29	36.71
	China	14	17.72
	India	12	15.19
	Others (Afghanistan, Bangladesh, Barbados, Colombia, Egypt, Israel, Lebanon, Oman, Pakistan, Palestine, Portugal, Saudi Arabia, USA)	20	25.35
	Prefer not to say	4	5.06

*Some columns do not total 100% due to rounding.

task materials and prioritize a backlog of user stories for a mall's digital screen display with facial recognition software, as explained in Section III-B. This task took approximately 50 minutes to complete.

E. Data Analysis

All collected data were double-entered into a spreadsheet by the first and second authors. Disagreements were found for 95 rows (4% of the data) and corrected by consensus by double checking the original data with both authors present.

Each row is a data point where a participant has prioritized a particular story type. Columns include participant ID, user story ID, story type, treatment group, priority assigned, gender, age, and country. Story type is a categorical variable with three possible levels: prosocial, neutral and antisocial. Treatment group is a categorical variable with four possible values: persona, stakeholder map, both, and control. Lastly, the prioritization level is an ordinal variable with values from 1 to 4.

Our analysis began by examining whether our dataset satisfies assumptions for Cumulative Link Mixed Model (CLMM), an extension of Ordinal Logistic Regression (OLR) (see section IV-A). CLMM is used to predict an ordinal dependent variable given one or more independent variables, accounting for fixed and random effects. CLMM is necessary because each participant prioritizes multiple stories, making the observations not wholly independent.

Next, we split the dataset and fit two CLMM models, one predicting the priorities assigned to prosocial stories; the other predicting the categories assigned to antisocial stories. We took a dual model approach for two reasons. First, our hypotheses focus on the interactions between the type of user story (prosocial, neutral, or antisocial) and the intervention groups (stakeholder map, personas, both, or neither). A single model would be more complicated and difficult to understand because it includes main effects (which are irrelevant to our hypotheses); neutral stories (also irrelevant to our hypotheses), and many more coefficients. Second, fitting a single model would lead to violations of CLMM's proportional odds assumption. Thus, we tested our hypotheses using two separate CLMM models.

For both models, the dependent variable is the prioritization level, which is ordinal. We recoded the categorical variable of the intervention group to binary variables [63], which allowed us to set the reference category by the order of independent binary variables. We set the control group as the reference category; that is, the category of the variable to which the other levels of the categorical variable are compared.

We fit the model using the ordinal package [64] for the R programming language (Version 4.1.2) [65].

Recall that we hypothesized that the stakeholder map and personas would lead to higher prioritization of prosocial user stories and lower prioritization of antisocial user stories. A significant negative coefficient for any of the predictors 'persona', 'stakeholdermap', and 'stakeholdermap * persona' with respect to the story type would indicate that the use of personas or stakeholder map or both effectively influences participants to prioritize antisocial stories lower or prosocial stories higher. Predictors with p-values less than 0.025 will be considered statistically significant.

Detailed analysis scripts are available in our comprehensive replication package (see Section VI)

IV. RESULTS

In this section, we report data diagnostics and hypothesis tests, and visualize our main results.

A. Data Diagnostics

CLMM makes four main assumptions:

1) The dependent variable is ordinal. Our dependent variable, Prioritization_Level, is in fact ordinal.

2) The independent variable is categorical. We recoded our independent variables to make them binary [63], which satisfies this assumption.

3) No multicollinearity among independent variables. We checked for multicollinearity with the vif function from

the `car` package [66]. All variance inflation factors were between 1.42 and 1.53, well below the threshold of 5 [67].

4) The effect of each independent variable is consistent across different thresholds of the ordinal dependent variable (Proportional Odds). We used the function `nominal_test` from the `ordinal` package to check for proportional odds [64]. The data was split by story type as described in Section III-E. The results show high p-values ($p > 0.05$) for 'stakeholder map' ($p = 0.23, 0.12$), 'persona' ($p = 0.69, 0.58$) and their interaction ($p = 0.072, 0.068$), indicating that the Proportional Odds assumption is met.

B. Hypothesis Tests

We test each hypothesis below using CLMM regression.

Hypothesis H1A: Priorities assigned to prosocial user stories are higher when a stakeholder map is presented. With a positive coefficient of 0.3124 and a p-value 0.183, we fail to reject the null hypothesis (see Table III). The effect of the stakeholder map on prioritization of prosocial stories is not statistically significant.

Hypothesis H1B: Priorities assigned to antisocial user stories are lower when a stakeholder map is presented. With a coefficient of -0.2059 with a p-value of 0.5135, we fail to reject the null hypothesis (see Table IV). The effect of the stakeholder map on prioritization of antisocial stories is not statistically significant.

Hypothesis H2A: Priorities assigned to prosocial user stories are higher when a set of persona models is presented. With a positive coefficient of 0.1645 and p-value 0.501, we fail to reject the null hypothesis (see Table III). The effect of personas on prioritization of prosocial user stories is not statistically significant.

Hypothesis H2B: Priorities assigned to antisocial user stories are lower when a set of persona models is presented. With a significant negative coefficient of -0.8672 ($p = 0.009$), we reject the null hypothesis. The effect of personas on prioritization of antisocial stories is statistically significant (see Table IV). The corresponding odds ratio (0.42) indicates that the presence of personas reduces the odds of assigning a higher prioritization level to antisocial user stories by 58%.

Interaction effects. If there were a synergistic interaction between the treatments (i.e., stakeholder maps and personas are more effective together than individually) then absolute value of the coefficients (and odds ratios) of the experimental group receiving both treatments should exceed the absolute value of the coefficients (and odds ratios) of the experimental groups that received either treatment alone. This is not the case, so there is no evidence of synergistic interaction.

C. Data Visualization

Fig. 4 and Table V show mean prioritization scores across different intervention groups and user story types. The negative slopes of the lines indicate that participants in all groups prioritized prosocial stories higher than neutral stories and neutral stories higher than antisocial stories, on average.

The distances between the dashed red lines and solid green lines correspond to our hypothesis tests. In the leftmost plot, we see that participants gave similar average priorities to prosocial and neutral stories regardless of whether they received persona models. However, participants who received persona models assigned significantly lower priorities to antisocial user stories.

The middle plot shows how participants who received the stakeholder map gave slightly higher priorities to prosocial stories, and slightly lower scores to antisocial stories compared to the control group, but neither difference is statistically significant.

The rightmost plot shows how participants who received both the stakeholder map and personas gave somewhat lower scores to antisocial user stories, but this difference is not significant. Moreover, because the "both" group displays a smaller effect than the persona-only group, we see no visual evidence of synergistic interaction between the treatments.

V. Discussion

We hypothesized that exposure to a stakeholder map and persona models would lead participants to prioritize prosocial user stories higher and antisocial user stories lower. We found that the effect of personas is imbalanced: personas significantly lower priorities of antisocial stories but do not increase priorities of prosocial stories.

Stakeholder maps appear to display a more balanced pattern (boosting prosocial story priorities while dampening antisocial story priorities); however, these effects are not statistically significant, either because the pattern is just random noise or because the effect is too small to detect with the study's present sample size.

Similarly, we found no evidence of synergistic (or antagonistic) interaction between treatments. Again, this may be because the treatments do not interact or because the effect of stakeholder maps is too small to detect with a study of this size.

The fact that personas have a stronger effect is consistent with our proposed generative mechanism. Personas should affect prioritization by helping professionals empathize with users. More diverse personas, as used in this study, should help professionals empathize with a greater variety of users. In this sense, personas, with their pictures and details, are a more powerful trigger than stakeholder maps.

TABLE III
CUMULATIVE LINK MIXED MODEL (CLMM) RESULTS FOR PROSOCIAL USER STORIES PRIORITIZATION

Random Intercept for:	Participant.ID				
Intercept Variance	0.1475				
Intercept Std.Dev.	0.3840				
Number of Participants	79				
	Coefficients	**Std. Error**	**z value**	**P > |z|**	**odds ratio**
Reference Category: Control					
Stakeholdermap	0.3124	0.2344	1.333	0.1830	1.3666
Persona	0.1645	0.2442	0.674	0.5010	1.1787
Both	−0.0585	0.2163	−0.270	0.7870	0.9432

TABLE IV
CUMULATIVE LINK MIXED MODEL (CLMM) RESULTS FOR ANTISOCIAL USER STORIES PRIORITIZATION

Random Intercept for:	Participant.ID				
Intercept Variance	0.5759				
Intercept Std.Dev.	0.7589				
Number of Participants	79				
	Coefficients	**Std. Error**	**z value**	**P > |z|**	**odds ratio**
Reference Category: Control					
Stakeholdermap	−0.2059	0.3152	−0.653	0.5135	0.8139
Persona	−0.8672	0.3314	−2.617	0.0088 **	0.4201
Both	−0.5258	0.2966	−1.773	0.0762	0.5911
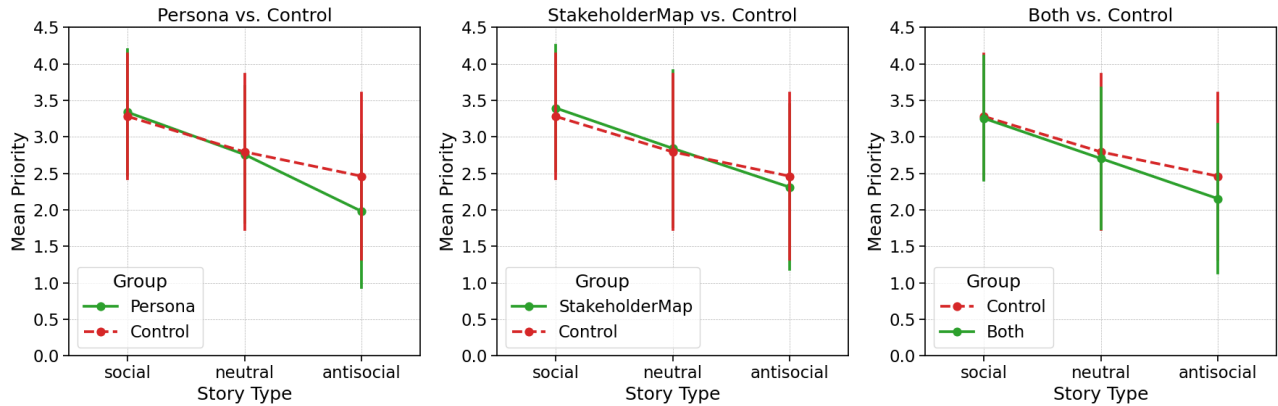
Fig. 4. Interaction Plots

TABLE V
MEAN PRIORITIZATION SCORES PER GROUP AND USER STORY TYPE

	antisocial	**neutral**	**social**	**n**
Stakeholder map	2.30	2.84	3.39	19
Persona	1.98	2.74	3.34	16
Both	2.15	2.70	3.28	24
Control	2.46	2.79	3.28	20

A. Implications

A recent scoping review found a lack of studies evaluating sustainability-improving interventions [2]. Meanwhile, Becker [68] emphasized the importance of enacting meaningful change in computing to address pressing sustainability issues. These works inspired the study described in this paper, which investigates specific interventions that could improve social sustainability in software development.

Based on the results reported above, we recommend that software teams consider incorporating personas into their software development process to promote socially sustainable development. Although personas only affected the prioritization of antisocial stories in our experiment, practically speaking, priorities are relative, so it may not matter much if an intervention boosts prosocial priorities or dampens antisocial priorities or a bit of both.

For researchers, this study contributes to the growing body of knowledge regarding the intersection of software engineering and social sustainability. To date, there is a lack of empirical (particularly experimental) validation of specific social

sustainability interventions, making our findings particularly noteworthy.

Furthermore, we suspect that the lack of experiments evaluating social sustainability interventions is due, in part, to the difficulty of assessing these interventions. We hope that other researchers will adopt our backlog-prioritization approach to assessing social sustainability interventions and provide all of our materials for this purpose (see Data Availability). A common assessment approach is more efficient and should improve replicability and validity (as different research teams refine and evolve one approach instead of re-inventing new ones).

For educators, our results support teaching persona modeling in the software engineering curriculum. Furthermore, we encourage educators to teach sustainable development in general and social sustainability in particular throughout computer science and engineering curricula. Simply promoting sustainability awareness and normalizing discussions thereof is half the battle.

B. Limitations

In this section we address standard quality criteria for experiments with human participants [69]. Like many randomized controlled experiments, this study sacrifices generalizability to maximize internal validity.

Internal validity is high because: (1) we used a well-understood, widely-used, fully-crossed, two-factor between subjects design in which participants were randomly assigned to experimental groups; (2) participants were unaware of the specific research focus (reducing social desirability bias); (3) each group was in its own room, reducing social threats to internal validity; (4) all participants were provided with the same prioritization worksheet and placed in similar experiment rooms with uniform experimental setups.

Data were double-entered to improve reliability and the high level of agreement achieved demonstrates good reliability. Conclusion validity is similarly high as we used a sophisticated regression technique (CLMM) that maximizes statistical power given constraints imposed by data types and distributions. We systematically identified all the assumptions of CLMM and verified that they are met). However, our sample size of 79 fell short of the suggested sample size of 104, inflating the probability of a type II error.

However, strong internal validity, reliability and conclusion validity come at the cost of poor external validity. We used a convenience sample of participants. While multicultural, the participants are all students in the same university so they may have other unknown similarities. Participants were students, not professionals, further threatening external validity (see [70], [71]). We used a single case (shopping mall facial recognition), a single stakeholder map, and a specific set of personas. Determining how sensitive our results are to all these specific details is not possible with this kind of study. A dif-

ferent sample of participants analyzing a different case using different task materials may act differently. While serious, these threats to internal validity are endemic to experiments with human participants; the whole point of an experiment is to maximize internal validity, not external validity [72].

Meanwhile, the construct validity of our study is nuanced. The priorities assigned by participants to stories are measured directly, not reflectively, so a psychometric approach to construct validity (e.g. convergent and discriminant validity) does not apply. We consulted with several industry experts to to ensure the priority levels reflect contemporary agile practice, and the prioritization categories were clearly explained to all participants in the task material and in the video shown at the beginning of the study. However, the reader is justified in wondering to what extent these priority ratings relate to social sustainability. In the broadest sense, pro- and antisocial simply mean good or bad for society, but there are no authoritative lists of what exactly are good or bad for society, and reasonable people both within and across cultures may disagree on whether a specific feature is good or bad for society.

The validity of our prioritization-based measurement approach therefore rests on our categorization of stories into prosocial, neutral, and antisocial. As explained in Section III-B, we conducted a series of pilot studies to ensure high agreement about the nature of each story. This approach is not perfect. It assumes that participants in the pilot study can estimate the social impact of the proposed user stories given the source materials and that the user stories can be neatly classified. Not every feature of every real software product can be neatly binned into prosocial, antisocial, or neutral, so maximizing agreement likely excludes more ambiguous stories. Furthermore, the pilot participants are not a representative sample of all of humanity. Nevertheless, the perfect agreement reached through the pilot process suggests good face validity. Furthermore, the fact that control group participants on average prioritized prosocial stories significantly higher than neutral stories and neutral stories significantly higher than antisocial stories suggests good predictive validity. In summary, then, since convergent and discriminant validity do not apply, we took reasonable steps to assess face and predictive validity. However, the imperfect nature of our pilot studies and oversimplification of features' social effects threaten construct validity in unquantifiable ways. Personas may be less effective at promoting social sustainability in projects full of controversial features that could be perceived as either pro- or antisocial.

C. Proposed Research Methodology for Social Sustainability Experiments

The lack of experimental evaluations of social sustainability-promoting interventions may be due in part to the methodological challenges of constructing such studies. Studies of software development processes (and modifications thereof) resist experimental closure [73]. That is, it is not possible to recruit hundreds of real software teams, randomly assign them to multiple groups, get one group to use some intervention

while another does not, then measure the sustainability of the systems they produce.

We therefore designed an experimental protocol for assessing social sustainability interventions to overcome four methodological challenges:

1) Expressed attitudes do not generally correspond to revealed attitudes [74], due to social desirability bias among other psychological factors [75]. The experimental task therefore should involve observable actions beyond expressing attitudes toward sustainability.

2) Most software professionals should be able to complete the experimental task. Therefore, we need a familiar, everyday context that does not require specialized domain knowledge (e.g. a web browser, not air traffic control software).

3) The methodology should be implementable as either an in-person, lab-based task (suitable for student participants) or a remote, web-based task (suitable for recruiting a geographically diverse sample of professionals).

4) The methodology should be easily adaptable to many different social sustainability promoting interventions, both to improve comparability across studies and to reduce effort associated with study design.

With these goals in mind, we formulated the method described in Section III. Developing the task materials took many iterations—significantly more effort than previous experiments with human participants we have conducted—but the resulting method appears to achieve all of our goals.

The task of prioritizing a backlog of user stories provides a good compromise. It is something software professionals do in the real world [76] that is easy to simulate and can give us insight into propensities (not just attitudes). When prioritizing the stories, participants consider many factors beyond social sustainability, so the tasks give us clues as to how much social sustainability considerations might affect participants' behaviour.

Meanwhile, we chose a facial recognition system because it is so ethically fraught that it is easy to generate clearly antisocial features. We chose a shopping mall setting because most people are familiar with shopping malls—shopping malls are ubiquitous throughout developed and middle-income countries and while they are concentrated in urban areas, so are software professionals. We pilot-tested our user stories with a culturally diverse group to reduce the likelihood that our pro- and anti-social categorization is culturally-relative.

While we implemented the study on paper, it is conducive to an online presentation. An interface for collecting story priorities would be trivial to create with any common questionnaire survey software. The only important difference is the ability to spread materials across a table. Interventions involving copious information may be more difficult to implement online.

In summary, we advance a novel approach for assessing social sustainability interventions in which experimental or simulation participants prioritize a backlog of prosocial, antisocial, and neutral user stories, and researchers compare mean priority ratings. We provide an easy-to-follow experimental protocol (Section III) and materials (see Data Availability) for other researchers to repeat the approach. Finding a reasonable way of assessing the impact of sustainability interventions in lab studies is an important step for facilitating future research in this area (next).

D. Future Work

Several avenues of future work are evident, including:

- Replicating the present study with professionals

- Conducting methodologically similar studies testing different interventions (e.g. exposure to postgraduate education in ethics and sustainability; participatory design).

- Refining our evaluation methodology to include: (1) ethically murkier stories with mixtures of pro- and anti-social effects; (2) greater variety of pro- and anti-social effects.

More studies proposing and rigorously evaluating interventions for promoting environmental, economic, and technical sustainability are also sorely needed.

VI. CONCLUSION

Sustainability is not just about building technically robust or energy-efficient products. Sustainability is also about infusing human values, ethics, and social responsibility into software products. The human and social aspects of software engineering are often neglected; yet it is increasingly clear that social sustainability is crucial for building software that makes the world better instead of worse. Recent studies have called on software engineering researchers to propose and validate more sustainability-promoting interventions (and write fewer position papers) [2], [39]. Our study has shown the possibilities of validating and assessing these kind of tools, starting with a controlled lab setting with student participants.

This paper makes two main contributions:

1) The first empirical evidence that a common software development practice (persona modeling) can affect social sustainability (in this case, by discouraging the selection and prioritization of antisocial features).

2) We provide a replicable, adaptable, extendable experimental approach for evaluating social sustainability interventions in lab- or web-based settings.

In summary, our findings suggest that technology companies and software professionals who value social sustainability should try user persona modeling. While we cannot say for sure that our lab study with students will generalize to any particular organizational context, persona modeling has no known risks or drawbacks.

We hope this study inspires other researchers to create and experimentally assess other sustainability-promoting interventions.

DATA AVAILABILITY

We provide a detailed replication package including our pilot study materials, experiment task materials, instructional videos, deidentified dataset, data diagnostics and data analysis script (https://doi.org/10.5281/zenodo.10827123).

REFERENCES

[1] G. H. Brundtland and World Commission on Environment and Development, *Report of the World Commission on Environment and Development: Our Common Future.* Oxford: Oxford University Press, 1987.

[2] S. McGuire, E. Schultz, B. Ayoola, and P. Ralph, "Sustainability is stratified: Toward a better theory of sustainable software engineering," in *IEEE/ACM 45th International Conference on Software Engineering (ICSE).* Melbourne, Australia: IEEE, 2023, pp. 1996–2008.

[3] S. Naumann, M. Dick, E. Kern, and T. Johann, "The greensoft model: A reference model for green and sustainable software and its engineering," *Sustainable Computing: Informatics and Systems*, vol. 1, no. 4, pp. 294–304, 2011.

[4] M. Al Hinai and R. Chitchyan, "Social sustainability indicators for software: Initial review," in *Proceedings of the Third International Workshop on Requirements Engineering for Sustainable Systems (RE4SuSy 2014)*, Karlskrona, Sweden, 2014, pp. 21–27.

[5] B. Penzenstadler and H. Femmer, "A generic model for sustainability with process- and product-specific instances," in *Proceedings of the 2013 Workshop on Green in/by Software Engineering*, ser. GIBSE '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 3–8.

[6] J. L. Gustavsson and B. Penzenstadler, "Blinded by simplicity: Locating the social dimension in software development process literature," in *Proceedings of the 7th International Conference on ICT for Sustainability*, ser. ICT4S2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 116–127.

[7] G. A. García-Mireles, "Environmental sustainability in software process improvement: a systematic mapping study," in *Trends and Applications in Software Engineering*, J. Mejia, M. Muñoz, Á. Rocha, T. San Feliu, and A. Peña, Eds. Cham: Springer International Publishing, 2017, pp. 69–78.

[8] H. Anwar and D. Pfahl, "Towards greener software engineering using software analytics: A systematic mapping," in *2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA).* IEEE, 2017, pp. 157–166.

[9] N. Wolfram, P. Lago, and F. Osborne, "Sustainability in software engineering," in *2017 Sustainable Internet and ICT for Sustainability (SustainIT).* IEEE, 2017, pp. 1–7.

[10] A. C. Moises, A. Malucelli, and S. Reinehr, "Practices of energy consumption for sustainable software engineering," in *2018 Ninth International Green and Sustainable Computing Conference (IGSC).* IEEE, 2018, pp. 1–6.

[11] B. C. Mourão, L. Karita, and I. do Carmo Machado, "Green and sustainable software engineering - a systematic mapping study," in *Proceedings of the XVII Brazilian Symposium on Software Quality*, ser. SBQS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 121–130.

[12] C. Calero and M. Piattini, "Puzzling out software sustainability," *Sustainable Computing: Informatics and Systems*, vol. 16, pp. 117–124, 2017.

[13] L. Duboc, S. Betz, B. Penzenstadler, S. Akinli Kocak, R. Chitchyan, O. Leifler, J. Porras, N. Seyff, and C. C. Venters, "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)*, 2019, pp. 6–16.

[14] R. Chitchyan, C. Becker, S. Betz, L. Duboc, B. Penzenstadler, N. Seyff, and C. C. Venters, "Sustainability design in requirements engineering: State of practice," in *Proceedings of the 38th International Conference on Software Engineering Companion.* Association for Computing Machinery, 2016, p. 533–542.

[15] B. Penzenstadler, A. Raturi, D. Richardson, C. Calero, H. Femmer, and X. Franch, "Systematic mapping study on software engineering for sustainability (se4s)," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 2014, pp. 1–14.

[16] C. Marimuthu and K. Chandrasekaran, "Software engineering aspects of green and sustainable software: A systematic mapping study," in *Proceedings of the 10th Innovations in Software Engineering Conference*, ser. ISEC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 34–44.

[17] G. A. García-Mireles, M. Á. Moraga, F. García, C. Calero, and M. Piattini, "Interactions between environmental sustainability goals and software product quality: A mapping study," *Information and Software Technology*, vol. 95, pp. 108–129, 2018.

[18] B. Penzenstadler, H. Femmer, and D. Richardson, "Who is the advocate? stakeholders for sustainability," in *2013 2nd International workshop on green and sustainable software (GREENS).* IEEE, 2013, pp. 70–77.

[19] S. I. Majumdar, M. S. Rahman, and M. M. Rahman, "Thorny issues of stakeholder identification and prioritization in requirement engineering process," *IOSR J. Comput. Eng.(IOSR-JCE)*, vol. 15, no. 5, pp. 73–78, 2013.

[20] L. Nielsen, *Personas-User Focused Design.* Springer, 2019.

[21] S. Faily and J. Lyle, "Guidelines for integrating personas into software engineering tools," in *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*, 2013, pp. 69–74.

[22] J. Salminen, K. Guan, S.-G. Jung, and B. J. Jansen, "A survey of 15 years of data-driven persona development," *International Journal of Human–Computer Interaction*, vol. 37, no. 18, pp. 1685–1708, 2021.

[23] E. Siddall, C. Baibarac, A. Byrne, N. Byrne, A. Deasy, N. Flood, C. Goulding, S. O'Driscoll, N. Rabbitt, A. Sweeney *et al.*, "Personas as a user-centred design tool for the built environment," in *Proceedings of the Institution of Civil Engineers-Engineering Sustainability*, vol. 164, no. 1. Thomas Telford Ltd, 2011, pp. 59–69.

[24] P. Lago, S. A. Koçak, I. Crnkovic, and B. Penzenstadler, "Framing sustainability as a property of software quality," *Communications of the ACM*, vol. 58, no. 10, pp. 70–78, 2015.

[25] P. Lago, R. Verdecchia, N. Condori-Fernandez, E. Rahmadian, J. Sturm, T. van Nijnanten, R. Bosma, C. Debuysscher, and P. Ricardo, "Designing for sustainability: lessons learned from four industrial projects," in *Advances and New Trends in Environmental Informatics: Digital Twins for Sustainability.* Springer, 2021, pp. 3–18.

[26] F. Albertao, J. Xiao, C. Tian, Y. Lu, K. Q. Zhang, and C. Liu, "Measuring the sustainability performance of software projects," in *2010 IEEE 7th International Conference on E-Business Engineering.* IEEE, 2010, pp. 369–373.

[27] M. Al Hinai and R. Chitchyan, "Engineering requirements for social sustainability," in *ICT for Sustainability 2016.* Atlantis Press, 2016, pp. 79–88.

[28] J. Von Geibler, C. Liedtke, H. Wallbaum, and S. Schaller, "Accounting for the social dimension of sustainability: experiences from the biotechnology industry," *Business Strategy and the Environment*, vol. 15, no. 5, pp. 334–346, 2006.

[29] S. McKenzie, "Adult and vocational education for social sustainability: A new concept for tvet for sustainable development," in *Work, Learning and Sustainable Development*. Springer, 2009, pp. 177–186.

[30] M. Al Hinai, "Quantification of social sustainability in software," in *2014 IEEE 22nd International Requirements Engineering Conference (RE)*. IEEE, 2014, pp. 456–460.

[31] N. Haj Ahmad, L. Stigholt, and B. Penzenstadler, "Ai systems' negative social impacts and their potential factors," 2022.

[32] S. Oyedeji, R. Chitchyan, M. O. Adisa, and H. Shamshiri, "Integrating sustainability concerns into agile software development process," *arXiv preprint arXiv:2407.17426*, 2024.

[33] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*. PMLR, 2018, pp. 77–91.

[34] J. Sin, R. L. Franz, C. Munteanu, and B. Barbosa Neves, "Digital design marginalization: New perspectives on designing inclusive interfaces," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–11.

[35] E. J. Rose, A. Edenfield, R. Walton, L. Gonzales, A. S. McNair, T. Zhvotovska, N. Jones, G. I. G. de Mueller, and K. Moore, "Social justice in ux: Centering marginalized users," in *Proceedings of the 36th ACM International Conference on the Design of Communication*, 2018, pp. 1–2.

[36] J. Stilgoe, R. Owen, and P. Macnaghten, "Developing a framework for responsible innovation," in *The Ethics of Nanotechnology, Geoengineering, and Clean Energy*. Routledge, 2020, pp. 347–359.

[37] M. Steen, T. Timan, and I. van de Poel, "Responsible innovation, anticipation and responsiveness: case studies of algorithms in decision support in justice and security, and an exploration of potential, unintended, undesirable, higher-order effects," *AI and Ethics*, vol. 1, no. 4, pp. 501–515, 2021.

[38] M. Al Hinai and R. Chitchyan, "Building social sustainability into software: Case of equality," in *2015 IEEE Fifth International Workshop on Requirements Patterns (RePa)*. IEEE, 2015, pp. 32–38.

[39] A. C. Moises de Souza, D. Soares Cruzes, L. Jaccheri, and J. Krogstie, "Social sustainability approaches for software development: A systematic literature review," in *International Conference on Product-Focused Software Process Improvement*. Springer, 2023, pp. 478–494.

[40] C. Pacheco and I. Garcia, "A systematic literature review of stakeholder identification methods in requirements elicitation," *Journal of Systems and Software*, vol. 85, no. 9, pp. 2171–2181, 2012.

[41] T. Johann and W. Maalej, "The social dimension of sustainability in requirements engineering." in *RE4SuSy@ RE*, 2013.

[42] D. Wheeler, B. Colbert, and R. E. Freeman, "Focusing on value: Reconciling corporate social responsibility, sustainability and a stakeholder approach in a network world," *Journal of general management*, vol. 28, no. 3, pp. 1–28, 2003.

[43] C. Becker, S. Betz, R. Chitchyan, L. Duboc, S. M. Easterbrook, B. Penzenstadler, N. Seyff, and C. C. Venters, "Requirements: The key to sustainability," *IEEE Software*, vol. 33, no. 1, pp. 56–65, 2015.

[44] T. R. D. Saputri and S.-W. Lee, "Integrated framework for incorporating sustainability design in software engineering life-cycle: An empirical study," *Information and Software Technology*, vol. 129, p. 106407, 2021.

[45] A. Cooper, *The Inmates are Running the Asylum*. Wiesbaden: Vieweg+Teubner Verlag, 1999, pp. 17–17.

[46] T. Huynh, A. Madsen, S. McKagan, and E. Sayre, "Building personas from phenomenography: a method for user-centered design in education," *Information and Learning Sciences*, vol. 122, pp. 689–708, 2021.

[47] A. Dahiya and J. Kumar, "Effect of user information on conceptual design thinking: A linkographic study," *Journal of Design Thinking*, vol. 2, no. 2, pp. 97–112, 2021.

[48] D. Karolita, J. McIntosh, T. Kanij, J. Grundy, and H. O. Obie, "Use of personas in requirements engineering: A systematic mapping study," *Information and Software Technology*, p. 107264, 2023.

[49] C. G. Hill, M. Haag, A. Oleson, C. Mendez, N. Marsden, A. Sarma, and M. Burnett, "Gender-inclusiveness personas vs. stereotyping: Can we have it both ways?" in *Proceedings of the 2017 chi conference on human factors in computing systems*, 2017, pp. 6658–6671.

[50] J. Grudin and J. Pruitt, "Personas, participatory design and product development: An infrastructure for engagement," in *Proc. PDC*, vol. 2, 2002, pp. 144–152.

[51] J. Pruitt and J. Grudin, "Personas: practice and theory," in *Proceedings of the 2003 conference on Designing for user experiences*, 2003, pp. 1–15.

[52] B. J. Jansen, S.-G. Jung, L. Nielsen, K. W. Guan, and J. Salminen, "How to create personas: three persona creation methodologies with implications for practical employment," *Pacific Asia Journal of the Association for Information Systems*, vol. 14, no. 3, p. 1, 2022.

[53] P. Losana, J. W. Castro, X. Ferre, E. Villalba-Mora, and S. T. Acuña, "A systematic mapping study on integration proposals of the personas technique in agile methodologies," *Sensors*, vol. 21, no. 18, p. 6298, 2021.

[54] C. N. Chapman and R. P. Milham, "The personas' new clothes: methodological and practical arguments against a popular method," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 5. SAGE Publications Sage CA: Los Angeles, CA, 2006, pp. 634–636.

[55] T. Matthews, T. Judge, and S. Whittaker, "How do designers and user experience professionals actually perceive and use personas?" in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 1219–1228.

[56] C. Chiarcos, B. Claus, and M. Grabski, *Salience: Multidisciplinary perspectives on its function in discourse*. Walter de Gruyter, 2011, vol. 227.

[57] A. Pooresmaeili, D. R. Bach, and R. J. Dolan, "The effect of visual salience on memory-based choices," *Journal of neurophysiology*, vol. 111, no. 3, pp. 481–487, 2014.

[58] E. Peters, D. Västfjäll, T. Gärling, and P. Slovic, "Affect and decision making: A "hot" topic," *Journal of behavioral decision making*, vol. 19, no. 2, pp. 79–85, 2006.

[59] K. Power, "Stakeholder identification in agile software product development organizations: A model for understanding who and what really counts," in *Proceedings of the 2010 Agile Conference*, ser. AGILE '10. USA: IEEE Computer Society, 2010, p. 87–94.

[60] R. Newcombe, "From client to project stakeholders: a stakeholder mapping approach," *Construction management and economics*, vol. 21, no. 8, pp. 841–848, 2003.

[61] J. Beattie, J. Baron, J. C. Hershey, and M. D. Spranca, "Psychological determinants of decision attitude," *Journal of Behavioral Decision Making*, vol. 7, no. 2, pp. 129–144, 1994.

[62] E. Erdfelder, F. Faul, and A. Buchner, "Gpower: A general power analysis program," *Behavior research methods, instruments, & computers*, vol. 28, no. 1, pp. 1–11, 1996.

[63] M. Hardy and J. Reynolds, "Incorporating categorical information into regression models: The utility of dummy variables," *Handbook of data analysis*, pp. 229–255, 2004.

[64] R. H. B. Christensen, *ordinal—Regression Models for Ordinal Data*, 2019, r package version 2019.12-10. [Online]. Available: http://www.cran.r-project.org/package=ordinal/

[65] R Core Team, *R: A Language and Environment for Statistical Computing*, 2021, version 4.1.2. [Online]. Available: https://www.r-project.org/

[66] J. Fox, S. Weisberg, D. Adler, D. Bates, G. Baud-Bovy, S. Ellison, D. Firth, M. Friendly, G. Gorjanc, S. Graves *et al.*, "Package 'car'," *Vienna: R Foundation for Statistical Computing*, vol. 16, no. 332, p. 333, 2012.

[67] M. O. Akinwande, H. G. Dikko, A. Samson *et al.*, "Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis," *Open journal of statistics*, vol. 5, no. 07, p. 754, 2015.

[68] C. Becker, *Insolvent: How to Reorient Computing for Just Sustainability*. Boston, USA: MIT Press, 2023.

[69] P. Ralph, S. Baltes, D. Bianculli, Y. Dittrich, M. Felderer, R. Feldt, A. Filieri, C. A. Furia, D. Graziotin, P. He, R. Hoda, N. Juristo, B. A. Kitchenham, R. Robbes, D. Méndez, J. S. Molléri, D. Spinellis, M. Staron, K. Stol, D. A. Tamburri, M. Torchiano, C. Treude, B. Turhan, and S. Vegas, "ACM SIGSOFT empirical standards," *CoRR*, vol. abs/2010.03525, 2020. [Online]. Available: https://arxiv.org/abs/2010.03525

[70] I. Salman, A. T. Misirli, and N. Juristo, "Are students representatives of professionals in software engineering experiments?" in *2015 IEEE/ACM 37th IEEE international conference on software engineering*, vol. 1. IEEE, 2015, pp. 666–676.

[71] D. Falessi, N. Juristo, C. Wohlin, B. Turhan, J. Münch, A. Jedlitschka, and M. Oivo, "Empirical software engineering experts on the use of students and professionals in experiments," *Empirical Software Engineering*, vol. 23, pp. 452–489, 2018.

[72] K.-J. Stol and B. Fitzgerald, "The abc of software engineering research," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 27, no. 3, pp. 1–51, 2018.

[73] D. P. Ralph, "Fundamentals of software design science," Ph.D. dissertation, University of British Columbia, 2010.

[74] M. Fishbein and I. Ajzen, *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison Wesley, 1975.

[75] R. J. Fisher, "Social desirability bias and the validity of indirect questioning," *Journal of consumer research*, vol. 20, no. 2, pp. 303–315, 1993.

[76] T. Sedano, P. Ralph, and C. Péraire, "The product backlog," in *Proceedings of the 41st International Conference on Software Engineering*, ser. ICSE '19. IEEE Press, 2019, p. 200–211.