Diversity Awareness in Software Engineering Participant Research

Riya Dutta

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Software Engineering) at

Concordia University

Montréal, Québec, Canada

April 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Riya Dutta**

Entitled: **Diversity Awareness in Software Engineering Participant Research**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Software Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Jinqiu Yang

_____ Examiner
Dr. Marta Kersten-Oertel

_____ Examiner
Dr. Jinqiu Yang

_____ Co-Supervisor
Dr. Tanja Tajmel

_____ Co-Supervisor
Dr. Emad Shihab

Approved by   _____
                     Lata Narayanan, Chair
                     Department of Computer Science and Software Engineering

_____ 2023   _____
                              Mourad Debbabi, Dean
                              Faculty of Engineering and Computer Science

Abstract

Diversity Awareness in Software Engineering Participant Research

Riya Dutta

Diversity and inclusion are necessary prerequisites for shaping technological innovation that benefits society as a whole. A common indicator of diversity consideration is the representation of different social groups among software engineering (SE) researchers, developers, and students. However, this does not necessarily entail that diversity is considered in the SE research itself.

Our study examines how diversity is embedded in SE research, particularly research that involves participant studies. To this end, we selected 79 research papers containing 105 participant studies spanning three years of a renowned SE conference. Using a content analytical approach, we identified how SE researchers report the various diversity categories of their study participants and investigated: 1) the extent to which participants are described, 2) what diversity categories are commonly reported, and 3) the function diversity serves in the SE studies.

Our results demonstrate that even though most SE studies report on the diversity of participants, SE research often emphasizes professional diversity, such as occupation and work experience, over social diversity, such as gender or location of the participants. Furthermore, our results show that participant diversity is seldom analysed or reflected upon when SE researchers discuss their study results, outcomes, or limitations.

To help researchers self-assess their study diversity awareness, we propose a diversity awareness model and guidelines that SE researchers can apply to their research. Furthermore, our research contributes to the United Nations Sustainable Development's Goals 5, 9, and 10. With this study, we hope to shed light on a new approach to tackling the diversity and inclusion crisis in the SE field.

Related Publications

The following publications are related to this thesis.

- **Riya Dutta**, Diego Elias Costa, Emad Shihab and Tanja Tajmel. "Diversity Awareness in Software Engineering Participant Research". Accepted to 2023 IEEE/ACM International Conference on Software Engineering (ICSE)

Dedication

To my parents, friends, and all young women pursuing science

Acknowledgments

- With heartfelt gratitude, I extend my sincere thanks to my parents, Shobha and Satyajit Dutta, who have been the driving force behind my success and the reason I stand here today. Your unwavering love, understanding, support, and encouragement have been the foundation of my journey, and I am eternally grateful for having you both in my life.

- I would also like to express my utmost gratitude to my supervisors, Dr. Tanja Tajmel and Dr. Emad Shihab. Tanja, I am immensely grateful to you for recruiting me and inspiring me to pursue my Master's degree. Your encouragement and guidance have been instrumental in my journey. Emad, thank you for all the opportunities and continuous support. I would also like to extend my appreciation to Dr. Diego Elias Costa who has been an exceptional collaborator. It has been an honor to learn from each one you. Your invaluable insights and wisdom have made me a better researcher.

- My journey would not have been complete without the support and encouragement of the members of the DAS lab and the EDI lab. I extend my sincere thanks to all of you for making this experience more enjoyable. A special thank you to Mirjam and Ahmad for your great advice and encouragement.

- Lastly, I would like to thank my closest friends and confidants who have been my rock through thick and thin. Vincent, Diane and Michel, I am forever grateful for your unwavering support. To Lucas, Panorea, Jodie, Allison, Lori, Alvyn, Brianne, Richard, Daniel, Srishti and Momo, thank you for being my biggest fans and supporters. Your love, encouragement, and unwavering support have kept me motivated and inspired throughout this journey.

Preface

The lack of women in the field of engineering is a widely recognized issue. My journey to realizing this issue began when I started my Bachelor's degree in Software Engineering (SE) at Concordia University in August 2016. Coming from an all-girls school in India, I had never experienced the absence of female peers in my personal or academic life. However, upon joining university, I quickly became aware of the scarcity of women in my classes.

To fill this void, I joined an undergraduate association called Women in Engineering (WIE) at Concordia in 2018. This association not only provided me with a community of like-minded women in the field, but also opened my eyes to the various challenges faced by women in engineering on a daily basis. As an immigrant woman of color studying engineering in Canada, I embodied the intersectional identity of multiple minority groups as I not only represented a minority in terms of gender but also in other aspects such as race, nationality, and ethnic background. I was starting to recognize the societal challenges due to my intersectional identity.

It was during the last semester of my Bachelor's degree in Winter 2020 that I discovered the concepts of Equity, Diversity, and Inclusion (EDI) through a course called "Impact of Technology on Society." This course changed my perspective on engineering, allowing me to see it as not just a technical field, but also one that desperately requires attention to social aspects.

It was not in my original plan to pursue a Master's degree, and the prospect of interdisciplinary research combining SE and EDI was something I had never even imagined. However, the idea of integration of my technical background in engineering with my passion for EDI sparked my interest in pursuing a Master's degree. Moreover, as technology becomes increasingly ingrained in our daily lives, the importance of considering EDI in the field of SE has become increasingly relevant.

This project has not only provided me with invaluable research experience, but it has also opened my eyes to the stark differences in diversity considerations between different disciplines. When I first presented my research idea in a Master's class, many students from other fields were shocked to hear about the lack of social diversity considerations in the realm of Software Engineering.

Conducting research on diversity in an open-ended manner, as we did in this project, allowed me to broaden my understanding of diversity beyond just age, gender, and sex and allowed me to critically examine the concepts of diversity. I hope that my research can inspire other researchers to expand their views on diversity in a similar manner.

With my Master's research, I aim to embody my identity as both a software engineering researcher and a minority in society. I intend to carry the lessons I have learned from this thesis with me and strive to consider EDI in all aspects of my future, whether it be industry or academia. I am hopeful that in the future, more young women will pursue higher education in engineering.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis Introduction

In recent years, the consideration of the diversity of people who are involved in research and innovation has become an increasingly important topic. Diversity of developers and researchers is regarded – and has already been proven - beneficial for numerous reasons: diverse teams produce better outcomes (Menezes & Prikladnicki, 2018; Patrick & Kumar, 2012; Pieterse, Kourie, & Sonnekus, 2006), research and development that systematically includes diversity are of benefit for a broader population whereas the lack of diversity might result in biased and discriminatory technologies (Buolamwini & Gebru, 2018; Schiebinger et al., 2011-2020; Tannenbaum, Ellis, Eyssel, Zou, & Schiebinger, 2019), the inclusion of diverse stakeholders is paramount for the creation of an ethical and social-responsible future (Sarewitz, 2005; United Nations Department of Economic and Social Affairs, 2015; Van Oudheusden & Shelley-Egan, 2021), and, last but not least, excluding large parts of society from research and the development of future technologies violates human rights (Al-Nashif, 2021; Oberleitner, 2021; PAR-L, 2010). One size does not fit all, and this applies to research and development, too.

To mitigate the negative effects of non-diverse research and to stimulate researchers to incorporate diversity in their research, national and international research funding programs are developing policies and guidelines for equity, diversity, and inclusion (EDI) (NSERC, 2022), for responsible

research and innovation (RRI) (Parliament & Council, 2013) to bind the research funding to increasingly important conditions: the sufficient consideration of EDI.

In our study, we focus on research that includes research participants as we regard this as an evident opportunity for SE researchers to include diversity considerations in their research. Therefore, our research questions are as follows:

- RQ1. To what extent do SE participant studies report the diversity of participants?

- RQ2. What diversity categories are reported in SE participant studies?

- RQ3. What is the function of diversity in SE participant studies?

To capture all variants of diversity categories, we chose an inductive approach, which means: we do not predefine the categories to examine the content according to these categories. Instead, we reconstruct the diversity categories through open coding of the content. Therefore, for the purpose of this study, we chose a broader definition of diversity categories: A diversity category is a category based on which individuals (in our case, research participants) are distinguished and grouped. This open approach allows for identifying categories that might be relevant for research but are not included in the commonly discussed diversity categories such as gender, race, and ethnicity, amongst others (Nkomo & Stewart, 2006).

Our results demonstrate that only a few studies do not consider diversity at all, however, the examined studies differ greatly in the range of the consideration and reporting of diversity. From these outcomes, we draw the conclusion of differences in diversity awareness among SE researchers. Additionally, we propose a model of diversity awareness for participant studies as a tool to support SE researchers in reflecting on diversity and incorporating it systematically in their research. Finally, we publish our classification coding scheme and the dataset used to conduct this study to encourage further studies on diversity awareness and facilitate replicability[1].

1.2 Thesis Overview

This thesis is organized as follows:

---

[1]https://doi.org/10.5281/zenodo.7587076

- Chapter 2 describes the related literature. In this chapter we start by discussing diversity and intersectionality to establish a solid foundation for our thesis. Then we discuss sampling and representation because we examine participant studies in this research. We then delve into the literature related to diversity in SE. Finally, we examine research closely related to ours examining human aspects, such as diversity, within the SE research community.

- We detail our study methodology in Chapter 3 where we start with an overview of the research where we describe the mixed method approach we use for our research. We then define our three research questions. Finally, we describe in detail how we collect the sample data, define the coding scheme, and perform the content analysis.

- We present the results of our three research questions in Chapter 4.

- Upon reflecting on the results, we discuss a model of diversity awareness that researchers can use to self-assess their diversity awareness levels in Chapter 5. We also provide some guiding questions that can be used by researchers when conducting their research.

- We discuss the limitations of our study in Chapter 6 where we examine the internal, external and construct validity of our research.

- Finally, we present our conclusion, the impact this research has on society at large as well as the potential for future work in Chapter 7.

1.3   Thesis Contribution

This thesis presents a valuable contribution to researchers, the SE community, and society as a whole. In particular, it:

- Evaluates the current level of diversity consideration in SE research, specifically in regards to participant studies. This valuable information provides insight into the current state of the field and can be used to drive positive change towards a more diverse and inclusive approach to research.

- Presents a model and guiding questions for SE researchers to assess their own level of diversity awareness while conducting participant research. This provides a practical tool for researchers to ensure that they are taking diversity and inclusivity into account in their work.

- Provides the classification coding scheme and dataset used in conducting the study. This information encourages further research on diversity awareness and makes it possible for other researchers to replicate the study, adding to the body of knowledge in the field.

- Aligns with the United Nations Sustainable Development Goals 5, 9, and 10, which focus on gender equality, innovation and infrastructure, and reduced inequalities, respectively. By promoting diversity and inclusivity in SE research, this thesis makes a valuable contribution to achieving these important goals.

Chapter 2

Literature Review

In Chapter 2, we describe the related literature to our research taking a funnel approach where we detail important concepts and general related research before diving into more specific topics. We start off by providing a discourse on diversity and defining diversity to provide a solid basis for our research. We also define intersectionality as it is an important concept to consider when discussing diversity. Then we delve into understanding sampling and representation as this research focuses on participant studies. We proceed to investigate the issues related to sampling such as the I-methodology. As this is a SE research project, we examine research papers that are more closely related to ours which investigate the considerations of human values, such as diversity, in SE research. Finally, we will discuss the issue of the lack of diversity in SE by examining literature related to the ongoing diversity discussions in the SE world and demonstrate the gap our study fills.

2.1   Diversity Discourse and Intersectionality

Diversity has been long discussed in various fields in different contexts. The discourse on diversity originated in natural sciences and philosophy, where different diversity categories were created based on observed similarities between organisms (Litvin, 1997). However, in this thesis, we aim to contribute to the discourse on human diversity defined by social differences.

Diversity is a multifaceted term that can have various interpretations. Governments and organizations around the world have attempted to define diversity, each with their own unique perspectives.

The Canadian government, for example, defines diversity as encompassing "culture, ethnicity, religion, sex, gender, sexual orientation, age, language, education, ability, family status or socioeconomic status" (Committee et al., 2021).

The UN Declaration of Human Rights, Article 2 (Assembly et al., 1948) states, "Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status" (p. 6). These distinctions such as race, sex, and language, among others, mentioned by the UN, represent observed discrimination based on these grounds that contribute to the ever-evolving definition of diversity and the need to reflect on the processes of how differences are constructed among human beings.

The concept of diversity has been widely studied and discussed by researchers from different fields. As a result, different definitions of diversity have emerged in various contexts (Nagappan, Zimmermann, & Bird, 2013; Tamtik, 2022; Zanoni & Janssens, 2004).

Tamtik (2022) explored the shift in outlook of diversity in Canadian post-secondary institutions and they drew parallels with the changing outlook of diversity in society. The study discovered that intersectionality has been a more recent focus in publications, whereas earlier, diversity was narrowly defined in terms of gender, language, and socioeconomic categories. Tamtik (2022) notes that understanding diversity solely in terms of gender, language, and socioeconomic categories limits the understanding of diversity.

In their study, Zanoni and Janssens (2004) applied a critical discourse analysis approach to analyse texts from interviews with human resource managers in order to define diversity. The study found that HR managers typically define diversity in two main ways. The first is in relation to the socio-demographic characteristics of groups of people, such as gender, culture, or disability. The second way involves defining people and grouping them based on their differences in abilities or performance in their organization. The researchers concluded that power relations play a significant role in defining diversity.

Nagappan et al. (2013) investigated whether SE researchers used diverse and representative project samples for their studies. They proposed a technique for SE researchers to assess the diversity and representativeness of their sample projects, which were drawn from technical tracks of two

years of ICSE and FSE conferences. This study focused on the diversity of projects rather than that of people.

In our research, **we adopt an inductive approach to defining diversity, reconstructing diversity categories based on the words and phrases used to characterize participants in our sample.** This approach offers a comprehensive understanding of the various criteria used to define people in software engineering research, as opposed to relying on predefined categories that may limit our understanding of diversity.

We recognize that "diversity categories" are not innate but are social differences constructed by society through historical contexts and power relations (Haslanger, 2012; Schelenz, 2022; Zanoni & Janssens, 2004). We are reflecting on these categories in our research and are contributing to a non-essentialist understanding of diversity (Hearn & Louvrier, 2015; Zanoni & Janssens, 2004).

The purpose of development of diversity categories is not necessarily to reinforce social differences but to recognize oppressed groups and hopefully address the imbalance of power in society. Vertovec (2012) addresses this when examining diversity in their publication stating, "The multiple purposes of different "diversity" initiatives roughly lie between anti-discrimination and positive acceptance" (p. 297). It is, therefore, essential to reflect on these diversity categories to acknowledge social differences and create more inclusive practices.

It is important to note that when considering the diversity of an individual, it is important not to adopt a uni-dimensional approach since people live multifaceted lives. For example, a person can belong to more than one diversity category, such as being a woman, who is disabled and belongs to an ethnic minority group. Therefore, making the consideration of the intersection of diversity categories or employing an intersectional approach is essential in research and innovation.

The term "intersectionality" was introduced by Kimberlé Crenshaw to describe how different types of discrimination, power, and privilege come together in the lives of Black women, and how this is not fully recognized when racism and sexism are considered independently (Crenshaw, 1989). The term has since then been used by many scholars to discuss the intersecting facets of privilege, domination or oppression faced by various groups of people (Collins & Bilge, 2020; Hankivsky, 2014; McKinzie & Richards, 2019). Collins and Bilge (2020) define **intersectionality** as follows:

"Intersectionality investigates how intersecting power relations influence social relations across diverse societies as well as individual experiences in everyday life. As an analytic tool, intersectionality views categories of race, class, gender, sexuality, class, nation, ability, ethnicity, and age — among others — as interrelated and mutually shaping one another. Intersectionality is a way of understanding and explaining complexity in the world, in people, and in human experiences" (p. 2).

A common concept often referred to when discussing intersectionality is the intersecting axes of privilege, domination and oppression (Pauly Morgan, 1996), that refers to the interconnected systems of power and inequality that shape individuals' experiences based on their social identities and characteristics. The intersecting axes form a sort of wheel divided into two halves, where the top half represents dominant groups that experience privilege and the bottom half represent the oppressed groups. Each axis represents a different category such as gender, sex, race, sexuality, ability, education, age, appearance, class, language, religion, etc. Therefore, we can interpret the wheel as such where, for example in the race category, white individuals hold power and privilege over the oppressed or marginalized groups of people of colour. The axes are interconnected, and this is important to note because individuals may face privilege, domination and oppression in different ways depending on the intersection of their various social categories they belong to.

Understanding the interconnected nature of the axes of privilege, domination, and oppression is important in creating a more equitable society. It requires acknowledging the ways in which systems of power and inequality are interconnected and working towards dismantling them to build a more diverse, representative, and inclusive society.

2.2   Sampling and Representation

An important question to consider for studies that involve research participants is how to select participants and how to define the sample. This section will give an overview of sampling methods to better understand how diversity can be considered in this process.

A research sample is most often defined as a group of people or participants selected from a larger population to collect data for research purposes (Acharya, Prakash, Saxena, & Nigam, 2013; Blackstone, 2018). The process of analysing a sample to make research observations is called

sampling (Blackstone, 2018). Sampling can be used to perform both qualitative and quantitative research. Qualitative researchers often apply nonprobability sampling methods to their research when conducting participant studies, according to Blackstone (2018). Nonprobability sampling is where all members of a population do not have the same likelihood of being selected as part of a research sample (Blackstone, 2018). Some common nonprobability sampling techniques are purposive, snowball, quota, and convenience sampling.

These four nonprobability sampling methods are defined as follows (Blackstone, 2018). Firstly, purposive sampling involves selecting participants based on specific criteria, such as their expertise or experience with a particular phenomenon. This technique is useful for studies where specific perspectives are sought. Secondly, snowball sampling involves asking participants to identify other individuals who may be relevant to the study, leading to a chain of referrals and the selection of additional participants. This technique can be useful for studies where participants are difficult to reach or identify. Thirdly, quota sampling involves selecting a sample that meets a predetermined quota for certain characteristics, such as age, gender, or ethnicity, rather than randomly selecting participants from the population. Fourthly, convenience sampling involves selecting participants who are readily available and easily accessible to the researcher, which can be useful for studies with limited time or resources. In all these sampling methods, representation of a sample with respect to the population is not a main consideration.

Unlike qualitative research, quantitative research often applies a probabilistic method of sampling. In probability sampling, the likelihood of a participant being selected as part of a sample is determined, as opposed to nonprobability sampling. Some common probability sampling methods are simple random, systematic, stratified and cluster sampling.

The following paragraph describes four probability sampling methods (Blackstone, 2018; Fricker, 2008). Simple random sampling is the first method and involves selecting participants randomly from a larger population, with each individual having an equal chance of being selected. This method can be done using random number generators or by drawing names from a hat. The second method, systematic sampling involves selecting participants at regular intervals from a larger population. For example, every 10th person on a list of a population could be selected. The third method, stratified sampling involves dividing a larger population into subgroups, or strata, based on

certain characteristics such as age, gender, or socioeconomic status. Participants are then randomly selected from each subgroup, with the goal of ensuring that each subgroup is represented in the sample. The fourth method, cluster sampling involves dividing a larger population into clusters, or groups, and then randomly selecting clusters to be included in the sample. Participants within the selected clusters are then included in the sample. The goal when employing probability sampling techniques is to identify a sample that is representative of the population from which the sample is selected.

A **representative sample** is a subset of a larger population that accurately reflects the characteristics and diversity of the entire population. This means that the sample should be chosen in a way that each member of the population has an equal likelihood of being selected, and the sample should be large enough to minimize sampling errors and biases. By using a representative sample, researchers can make valid inferences about the characteristics and behaviors of the larger population based on the data collected from the sample.

In software engineering research, representative samples are a rarity as assessed by Baltes and Ralph (2022). They investigate sampling in software engineering research by, applying a similar approach to ours, conducting a critical review of 115 research papers from a few renowned software engineering conferences. They provide the strengths and limitations of each sampling technique and provide guidance on how to choose the appropriate technique depending on the research. The authors emphasize that a sample is representative of a population if the relevant parameters correspond. Our results confirm that relevant parameters need to be considered to have a representative sample, additionally we discuss the social parameters of diversity and representation in our research.

Baltes and Ralph (2022) also state how the lack of representative samples undermines a scientific field. Representation of research participants in a sample is an important scientific consideration. If the sample is not representative of the population of interest in a systematic manner, then any analysis based on that sample will be biased (Fricker, 2008). For example, if a study aims to conclude something based on the differences of men and women, it is important to make sure that the sample does not comprise just men or just women. It is important to note here that representation goes beyond gender. The diverse characteristics of participants that need to be identified and considered depends on their relevance to the research itself.

Some of the most common diversity categories, as considered by the United Nations Statistics Division (n.d.), are divided in six main categories: i) basic demographic population data such as age, sex, number of children, rural/urban residence and marital status; ii) educational characteristics such as school attendance, literacy rate, and educational attainment; iii) ethnocultural characteristics such as national, ethnic, or religious groups; iv) migrant stock such as citizenship and foreign-born populations; v) household characteristics such as type of living quarters and head of household; vi) economic characteristics such as employment, occupation and industry. These categories mentioned by the United Nations Statistics Division (n.d.) represent census data collected from the populations of various countries and serve as just some examples of common diversity categories. There are many more diversity categories that can be considered in research, as explained in the upcoming sections of this thesis.

2.3   Issues Related to Sampling and I-methodology

While having a demographic representative sample is important, it is not always sufficient for certain research questions or situations. Demographic characteristics such as age, gender, and ethnicity can provide a broad understanding of a population, but they do not capture the complexity and diversity within each group, especially when considering diversity that is pertinent to a specific research topic. It is well known that certain groups of people, such as women, are underrepresented in the field of engineering (Blaisdell, 1994; Hill, Corbett, & St Rose, 2010; McIlwee & Robinson, 1992). It is important to include underrepresented groups when selecting participants for a study because this allows researchers to prevent the data collected from normative participants to overwhelm the data collected from participants belonging to underrepresented groups (Fernandez et al., 2016).

Only considering demographics and disregarding underrepresented groups in datasets can lead to amplifying bias in technology which could even result in threatening the basic civil rights of people. One such example of technological bias leading to discriminatory practices is the use of facial recognition software by law enforcement in America. Facial recognition systems have shown to misidentify people of colour, an underrepresented group, which could mean innocent individuals

11

can be subject to unwarranted police searches or wrongful arrests (Buolamwini & Gebru, 2018). The research by Garvie, Bedoya, and Frankle (2019) investigates the use of facial recognition software across 100 police departments in America revealing that the software inordinately affects African-Americans by making them most likely to be targeted by police and subject to searches. When an individual is arrested, the police take their mugshot which is then stored into a database which is often used as the training dataset for these facial recognition systems. Since people of colour are more likely to be wrongfully arrested, the mugshot database used to train the software is saturated with a disproportionately high number of African-Americans causing them to be subject to further bias(Garvie et al., 2019). This example shows how not considering the perpetual bias in society when collecting data for research can lead to algorithmic bias causing further discrimination towards certain groups.

Another example of algorithmic bias due to underrepresentation is presented by Lambrecht and Tucker (2018) where they observed discrimination presented by biased AI advertisement algorithms. They investigate an algorithm that was meant to advertise jobs for STEM fields in a gender-neutral way which revealed that women were less likely to see the ad than men, despite being more likely to click on it if they did see it. The study found that this gender imbalance did not correlate with World Bank measures of sexism or country wealth, but rather reflected the fact that women are a valuable target demographic and more expensive to advertise to. Consequently, an algorithm that aims to optimize ad delivery based on cost-effectiveness can unintentionally discriminate against women, even in the case of gender-neutral ads. These examples show how not considering the underrepresentation of certain groups or diverse populations in datasets and software algorithms can lead to exemplifying discriminatory and non-inclusive practices.

The limited consideration of diversity in research has resulted in proven problems. For example, the development of car crash test dummies were designed based on the average male body, leading to potential safety issues for people of different genders and body types (Schiebinger et al., 2011-2020). This issue arose because the researchers and creators of this innovation lacked diversity, which may have caused their inherent biases to impact the product, even subconsciously. Therefore, it is crucial to keep a diverse group of users in mind when developing new technologies or conducting research, and to be mindful of the potential biases that may arise from relying solely on

the "I-methodology". Oudshoorn, Rommes, and Stienstra (2004) describe the **I-methodology** as "a design practice in which designers consider themselves as representative of the users". It is when the designer of a product, most often unconsciously, embeds their own experience or perspective into creating a product, viewing themselves as the user (Akrich, 1992). An example of this issue is the case of "Digital cities" in the Netherlands, which were explored by Oudshoorn et al. (2004). Although the goal of these cities was to be accessible to everyone, the designers of the digital cities, who were mostly men with advanced technical knowledge, relied on the I-methodology, which led to the cities being primarily tailored to the needs of male users who were more tech savvy, making them not inclusive or accessible for users who were less tech savvy or female users. This case illustrates the importance of considering diversity in research and design, especially in fields like engineering that often lack diversity. By avoiding the I-methodology and prioritizing inclusivity, designers and researchers can create products that better serve the needs of all users, not just a select group.

This thesis addresses the issue of I-methodology by critically analysing diversity considerations in software engineering research. We propose a Diversity Awareness model and provide guidelines that researchers can use to assess their diversity awareness while conducting research. Considering diversity is essential for creating more representative, diverse, and inclusive research practices, as shown by previous sections of this thesis.

2.4   Human Values in Software Engineering Research

More recently, there has been a growing interest in studying the consideration of human values, such as diversity, and social aspects in SE research. Perera et al. (2020) examine the extent to which human values were considered in SE research. They use a similar methodology to ours, where they analyse research papers from top-tier SE conferences and journals to assess their consideration of human values. The found that very few SE publications consider human values.

Storey, Ernst, Williams, and Kalliamvakou (2020) conducted a similar study where they investigate the consideration of human aspects in SE research by analysing SE research papers. They

define human aspects as the human and social issues that affect human stakeholders, such ad developers, affected by research and development in SE. The study found that most SE studies focus on technical aspects, as opposed to human aspects, of SE even though these studies claim to impact human stakeholders. Storey et al. (2020) concluded that there is a need for SE studies to consider more human aspects and recommended a framework they created that can be used to consider more human and social aspects in SE research.

Although the work by Perera et al. (2020) and Storey et al. (2020) used a similar approach to ours, these two studies investigate the considerations of human aspects/values in SE research whereas we study the consideration of diversity in SE research, more specifically participant studies, while also providing guidelines for SE researchers conducting such studies.

The study that more closely aligns with ours is that of Lenarduzzi, Dieste, Fucci, and Vegas (2021), who investigated participant studies in SE by examining current participant selection guidelines and practices in empirical SE research. The researchers analysed existing guidelines for participant selection in SE and presented the participant selection strategies currently in use in their results. They investigated the participant selection strategies and identified validity threats related to these strategies in 118 software engineering studies. The majority of the studies (90 out of 118) selected participants based on their current availability, while some selected participants based on their experience in the study context or domain-specific skills. The authors of the studies typically only tested participants' experience in the domain via preliminary surveys for industrial participants. Only 50 out of the 118 studies mentioned possible limitations due to participant selection, with the most common issue being that selecting participants with different skills or expertise may have resulted in different outcomes. However, none of these studies discussed possible mitigation actions. Our study complements theirs by examining actual research papers that involve participants and providing guidelines for SE researchers, whereas their study only analysed guidelines (rather than actual research papers).

Additionally, the study by Schelenz (2022) closely relates to ours where they investigate the concepts of diversity in computer science and technology. The methodological approach applied by Schelenz (2022) is similar to ours, in which they "reconstruct diversity concepts from the literature by mapping their meaning and clustering similar meanings into categories" (p. 3). Both

our studies demonstrate a critical awareness of constructing difference by operating with categories. However, our sample differs in that we conduct our research by analyzing research papers with participant studies, whereas Schelenz (2022) analyzes research papers about diversity. Schelenz (2022) mentions how diversity-aware technologies are understudied, and our research fills this gap by investigating diversity awareness in software engineering research.

2.5   Lack of Diversity in Software Engineering

The lack of diversity in the field of SE is well known. In recent years, diversity, in particular, gender diversity, has garnered increasing attention in the field of SE (Bosu & Sultana, 2019; Burnett, Peters, Hill, & Elarief, 2016; Catolino, Palomba, Tamburri, Serebrenik, & Ferrucci, 2019; Padala et al., 2022). The lack of diversity in SE, also known as the diversity crisis, has been extensively investigated by numerous SE research papers, as evidenced by the IEEE special issue article by Albusays et al. (2021). The article clearly underscores the significance of diversity and inclusion in the field of SE, posing a crucial question: "What are the relevant diversity parameters we should consider when exploring software development practices and technology?" (Albusays et al., 2021) (p. 20). The issue highlights seven discussing diversity and inclusion in SE concluding that modern software practice place a high priority on addressing challenges pertaining to diversity and inclusion.

Moreover, Rodriguez, Nadri, and Nagappan (2021) highlight the need for SE research to consider additional perceived diversity aspects. Our research aims to address these challenges in the context of SE research by (i) examining how SE researchers address diversity in their studies and report on it in their publications, and (ii) providing SE researchers with guidelines to use when conducting participant studies.

In addition, Menezes and Prikladnicki (2018) conducted a literature review and semi-structured interviews to explore the impact of diversity on processes in software development. They concluded that many challenges still exist in making SE work environments more diverse. Our work complements their research as we investigate the extent to which SE research considers diversity, the categories of diversity used, and the role of diversity in participant studies.

Our study contributes to diversity research in SE with a novel diversity awareness model as

well as guidelines for researchers to consider diversity in SE research. This research has been supervised by professors from different disciplines, including SE and EDI studies, providing an interdisciplinary perspective to critically study diversity in SE research.

Chapter 3

Methodology

In Chapter 3, we present an overview of the study design, iterate over the research questions and finally describe in detail each step of the methodology used.

3.1 Overview

The objective of our research is to explore the extent to which diversity is acknowledged in SE research and to provide guidelines for SE researchers that they can use when conducting participant studies. To accomplish this, we employ a mixed-methods approach that combines qualitative and quantitative content analysis techniques (Krippendorff, 2018; Krippendorff & Bock, 2009). Our methodology involves using open coding and axial coding methods (Saldaña, 2021) to examine SE research papers that include participant studies.

Before embarking on our full study, we conducted a pilot study to evaluate the feasibility of our approach and to develop an appropriate framework. Figure 3.1 provides an overview of our methodology. First, we selected a venue for collecting data for our study. The pilot study was then carried out to assess the viability of our study and to establish a coding scheme. After the coding scheme was finalized, we expanded the data collection process to include a larger sample of studies with participants.

To gather our data, we meticulously screened abstracts of research papers from our selected
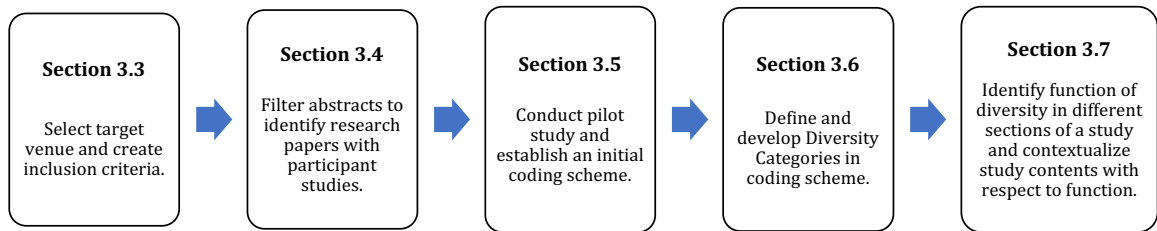
Figure 3.1: Overview of methodology

venues to identify studies that involved participants. I read each paper and classified each study according to the coding scheme. Agreement upon the coding scheme was met amongst three coders. Our approach leverages both qualitative and quantitative methods to provide a comprehensive analysis and reach our results.

3.2   Research Questions

To assess to what extent SE research papers with participants display diversity awareness, we first investigate if SE research papers report on diversity in their studies:

RQ1. To what extent do SE participant studies report the diversity of participants?

After examining whether diversity is even reported, we set out to identify the kinds of diversity reported:

RQ2. What diversity categories are reported in SE participant studies?

Finally, after assessing what kind of diversity categories can be identified in SE research with participants, we investigate if the reported diversity is actually analysed/leveraged throughout the research paper:

RQ3. What is the function of diversity in SE participant studies?

3.3 Selecting the Target Venue

To take a more systematic approach, we collected the sample for this research from the International Conference of Software Engineering proceedings (ICSE). We choose ICSE as our venue because it is considered the flagship software engineering conference (Damian & Zeller, 2022) and the proceedings of ICSE are easily accessible online (*ICSE '20: Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*, 2020). We collected our sample from ICSE 2019, 2020 and 2021 technical tracks to enrich the sample of studies and ensure our findings cover multiple editions of the conference. In total, the technical track of ICSE 2019, 2020, and 2021 contained 379 papers. Sampling three venues was also in line with software engineering meta-studies, such as the study of Storey et al. (2020), which used two venues, ICSE 2017 technical track and ESME 2017 paper, as their sample.

For completeness, we list below all of the inclusion criteria used for the data set of our study:

- Technical research track papers from ICSE 2019, 2020 and 2021.

- Studies that contained participants such as surveys, control groups, action research, grounded theory, focus groups, interviews, field studies, lab studies, validation studies, task-based studies and judgment studies.

- Studies with one or more participants

3.4 Selecting Papers with Participant Studies

Software engineering research employs a variety of research methods. As we are interested in studies with participants, we need to identify papers that have employed a particular set of methods, e.g., surveys, interviews, field studies, etc. To this aim, we read the abstracts of all 376 papers in the technical research track of ICSE, from 2019 to 2021.

We choose to read the abstracts of each paper in order to find research papers containing participant studies. Abstracts often convey the main methodology of a research, therefore if a participant

Table 3.1: The table shows some of the words and phrases we recognized in abstracts of papers containing participant studies.

Keywords indicating participant studies
judgement study, empirical studies, controlled experiment, control group, participants, analyze recordings, grounded theory methodology, grounded theory study, practitioners, surveys, surveyed, interviewed, collected quantitative information, focus group, action research, qualitative insights, quantitative survey, professional software developers, examined [...] developers, developers real work

study was conducted as part of a research, it would likely be mentioned in the abstract. This is particularly the case with participant studies, as the inclusion of interviews and surveys with practitioners is well-valued by the SE community; analysing paper abstracts is commonly used in related literature when conducting similar content analyses (Perera et al., 2020; Shaw, 2003). Given our initial sample contains 379 papers, reading the abstracts was a suitable and efficient method for filtering papers that were candidate for containing participant studies.

To select papers with participant studies, we identified words and phrases that indicated that the research involves research participants. Some examples of these words and phrases are given in Table. 3.1.

Sometimes, the abstract of a paper does not explicitly describe the entire methodology adopted by the studies. In the case of only implicit indication of research involving participants, we read the entire paper for verification. For example, in the following abstract, the indication of a study with participation is not clear.

"Using screencasts of developers' real work, we demonstrate the usefulness of our technique in a practical application for action-aware extraction of key-code frames in developers' work" (Zhao, Xing, Chen, Xia, & Li, 2019) (p. 350).

The highlighted text here indicates a possibility that the "developers" mentioned in the abstract are participants in this study. In such a case, we go beyond reading the abstract to verifying the contents of the paper to see if participants were actually involved in the study.

We did not consider studies where participants were not directly or actively involved in the research. For example, Wu *et al.* (Wu, Deng, Niu, & Nie, 2021) analysed app user reviews in their

research. Since the authors use already available reviews from the users and do not contact the app users directly to gather data from them, we do not consider this paper to have a participant study.

Table 3.2: The table shows the number of abstracts we read to select our sample, the number of research papers in our sample and the number of studies we found per venue.

	2019	2020	2021	Total
Number of Abstracts	109	129	138	376
Number of Research Papers	28	25	26	79
Number of Studies	34	33	38	105

Of a total of 376 papers from the three venues of ICSE 2019, 2020, and 2021 technical tracks, we identified 79 publications that met our selection criteria to be included in our study. A summary of this data is presented in Table 3.2. Out of our sample of 79 research papers with participants, some of the papers included more than one participant study. For example, Miller *et al.* (Miller, Rodeghero, Storey, Ford, & Zimmermann, 2021) conducted two different surveys with different sets of participants. Since the two participant studies that were described in the paper had different sets of participants, we coded the two different surveys as two different participant studies noting that they are part of the same research. Similarly, several other research papers consisted of two or more studies with participants. Therefore, even though we selected 79 research papers with participant studies as our sample data, we found a total of 105 studies with participants in those research papers. Thus, our main sample consists of a total of 105 participant studies.

3.5   Developing the Diversity Coding Scheme

To the best of our knowledge, there are no other studies that attempt to characterize diversity awareness of SE research. Hence, we cannot reuse an established framework to conduct our study. We start the study by developing an appropriate framework for encoding different diversity categories by conducting a pilot study using the ICSE 2020 technical research track. The ICSE 2020 technical research track contains a total of 129 papers. We read the abstracts of the 129 papers and identified 25 research papers with participant studies.

We chose the ICSE 2020 technical track for the pilot study instead of 2019 or 2021 tracks, because 2020 was the most recent track available during that time.

Once we selected our sample of 25 research papers, I read each research paper and identified references creating diversity categories and developing a coding scheme. By following an inductive approach, codes were developed while reading the papers. We used an open coding method where the codes emerged directly from the content of the paper (Saldaña, 2021). In the next step, we used the axial coding method where codes were grouped into diversity categories. Saldaña (2021), Tamtik (2022) and Storey et al. (2020) used a similar approach in their study. The coding scheme grew as new diversity categories were identified in the sample data.

Throughout the pilot study, the research team met weekly between February 2021 and June 2021 to discuss each paper in our sample. We discussed, defined and refined the diversity categories that emerged from each paper. The initial coding scheme developed through our pilot study was agreed upon by the research team. While the pilot study was fundamental for developing the initial coding scheme, the coding scheme was further refined and new codes were added to the coding scheme as we went through the rest of our sample. As the coding scheme evolved, all new codes added were discussed and agreed upon by the research team.

3.6   Qualitative Analysis of Papers to Create Diversity Categories

To address RQ2, we wanted to identify the various kinds of diversity being reported in SE participant studies, if at all. To this end, we used our coding scheme to code the sample of 105 studies. I read each paper and identified different words and phrases that describe and distinguish participants in the studies. For example, in the following quote from Krueger et al. (2020):

> "When participants elected to participate in the study, we collected basic demographic data (**sex**, **gender**, **age**, **cumulative GPA**, and **years of experience**) and **socioeconomic status** (SES) data" (p. 681).

the words highlighted in bold are descriptive words since they are used to describe the participants. We marked such descriptive words in each research paper.

We grouped codes to create the diversity categories. For example the words "sex", "gender", "gender fluid", "male", "female", "women", "sexual orientation" were all grouped into the diversity category labeled "gender/sex" . We met each week to discuss any potential ambiguities that arose in forming the diversity categories and find agreement.

These diversity categories were then added to our coding scheme. Whenever we discovered a descriptive word or phrase that could not previously be included in our coding scheme, we either added the new code to an existing category or added a new category to the coding scheme. Thus, our coding scheme included the diversity categories identified in our sample papers. It is important to note here that the diversity categories in our coding scheme are not all the diversity categories that exist. The ones in our scheme are just the ones we created based on the words and phrases we identified in our sample data. We wanted to inductively create diversity categories to add to our coding scheme rather than using pre-existing frameworks for diversity because we did not want to limit our view of diversity to the already existing diversity category definitions.

3.7 Identify the Function of Diversity Categories in Software Engineering Research

The goal of RQ3 is to understand how diversity is embedded in SE research. We want to identify the function that participant diversity serves in the selected SE studies. We define the "function" of diversity as the different aspects of diversity embedded in SE participant studies or the different roles that diversity plays in SE participant studies. We consider four different functions of diversity in participant studies:

(1) describing the participants of a study,

(2) analysing the impact of diversity in the study results,

(3) reflecting upon participant diversity in the study conclusion, and

(4) assessing the limitation of diversity in the participant study.

Given our sample contains 105 studies covering a multitude of different study types and goals, we evaluate the function of diversity in a study using a two-step approach. First, we identify the

23

section in which diversity is discussed, as the study section also serves a clear function in the study report. For example, we illustrate how a hypothetical ICSE paper would have its sections classified by our method in Fig. 3.2. Second, we carefully analyse the context in which diversity is inserted to confirm its function in the study.

In detail, we proceed as follows to identify the four functions of diversity:

- **Describing** the diversity of participants. This information is usually reported in the methodology sections of a research paper to describe the diverse sample of participants, e.g., in the "experimental design", "study design", and "research design" sections. In some studies, the diversity information of participants is not reported in the methodology section but the user evaluation part of their study. Hence the importance to consider the context in which diversity is described to identify its function in SE participant studies.

- **Analysing** the diversity of participants in study results. We evaluated the results reported in a research paper to identify if authors analyse the impact of participant diversity when discussing the results of their research. This information is usually reported in the "results" section of the paper but can also be found in "experimental results", "findings", etc.

- **Reflecting** upon the diversity of participants when concluding the study. We identified whether authors reflect on participant diversity and use this data to draw outcomes or conclusions in a paper. This information is typically discussed in the sections named "discussion", "conclusions" and "future work".

- Assessing the **limitations** of participant diversity in the study. To identify whether authors assess the limitations of participant diversity or the lack of collection of diversity data about participants, we evaluate the "threats to validity" or "limitations" section of research papers. It is important to identify whether authors assess the limitations of their participant diversity, as it reflects how aware authors are about diversity or lack thereof.

Each section of a paper serves a specific function, which is why we use the term 'function' in our study. Although the different sections of a paper may have different labels, as seen in the paragraphs above, their underlying purposes remain the same. Therefore, we choose to describe the
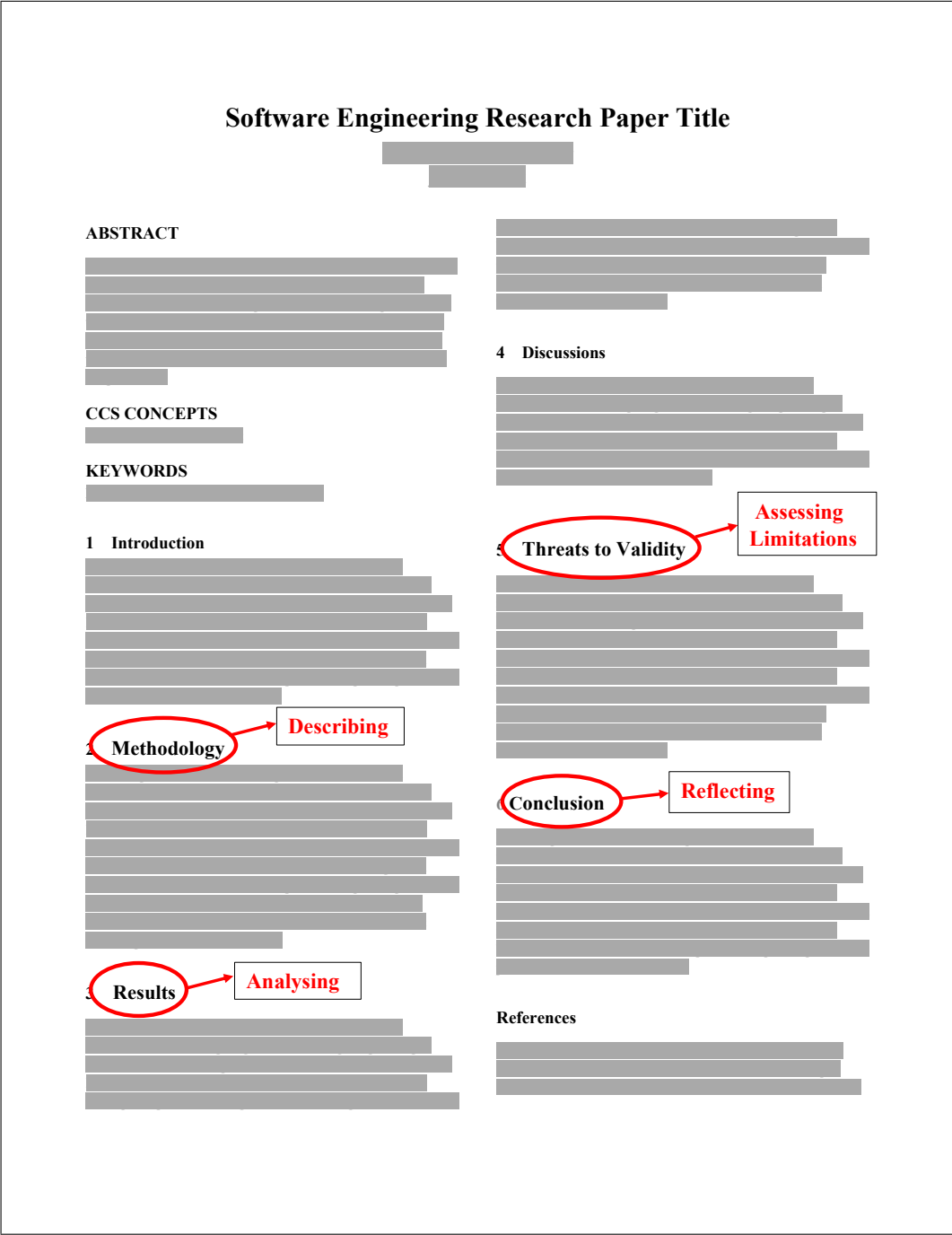
Figure 3.2: This image shows a generic layout format of a typical ICSE paper. The sections circled in red are the sections that typically correspond to the four functions of diversity, namely, describing, analysing, reflecting and assessing limitations in a study.

four different functions that diversity serves in a research paper, rather than referring to the specific section of the paper in which diversity is discussed.

Chapter 4

Results

The overall goal of our research was to assess the diversity awareness in software engineering research with participants. We felt it was necessary to shed light on the current state of diversity reporting in the field, in order to identify areas for improvement in the future. To achieve this, we posed three main research questions (RQ1, RQ2, and RQ3) to guide our investigation.

RQ1 aims to answer the question of whether diversity data about participants was being reported at all in software engineering studies. This was a fundamental question as it set the stage for our study and gave us a baseline understanding of the current state of diversity reporting.

RQ2 aims to uncover the type of diversity being reported in software engineering studies. This question helped us understand which aspects of diversity were being taken into consideration and reported in the field.

RQ3 aims to determine if diversity data was being reported only on a surface level or if it was being analysed further. This question was critical as it addressed the use of diversity data in software engineering research and provided insights into how it can be leveraged to drive positive change.

All of these research questions were designed to provide a comprehensive understanding of the level of diversity awareness among participants in software engineering research. The results for each research question are presented in depth in this chapter. Our findings provide valuable insights into the current state of diversity reporting in software engineering research and can serve as a foundation for future efforts to promote diversity and inclusion in the field.

4.1 RQ1. To what extent do SE participant studies report the diversity of participants?

The first step to understanding the awareness of diversity in SE research is to examine how often SE studies report the diversity of their participants. This analysis will help us understand the extent to which the practices of collecting, analysing and reporting diversity data about participants are well-established. Hence, to answer RQ1, we performed the qualitative content analysis described in Section 3.6. We also break our results by type of participant study (e.g., survey, task-based) and for this purpose we rely on how researchers describe their own study.

We found that 98.01% of the examined participant studies report the diversity of participants. Out of our sample of 105 studies that we analysed in our research, 103 studies used descriptive words and phrases when characterizing participants. Only 2 studies out of 105 did not report any participant diversity.

To make better sense of the results, we break down our analysis per type of participant study in Table 4.1. First, we note that SE papers use participant studies on 11 different study modalities, the most common being: surveys (51 studies), task-based (33 studies), interviews (22 studies), and validation (19 studies). Second, studies tend to report an average of 3 different categories when describing the participants, with laboratory or lab studies reporting the highest average (4.6 categories) across the samples we investigated. Some studies reported a wide array of diversity categories when describing participants, including 8 different types of descriptors.

Our sample included data on the number of participants reported by each study, as depicted in Figure 4.1. While we found no significant correlation between study size and the amount of diversity categories reported, we present the range of study sizes in our sample for completeness. Out of our sample of 105 studies, 5 studies failed to report the number of participants and therefore Figure 4.1 shows the number of participants reported by 100 studies.

Our results indicate that the practice of reporting participant diversity is well-established in SE participant studies. Researchers tend to select between 2-4 categories of descriptors when describing participants. Only in rare cases (2 out of 105 studies) do researchers omit to describe the participants included in their study.
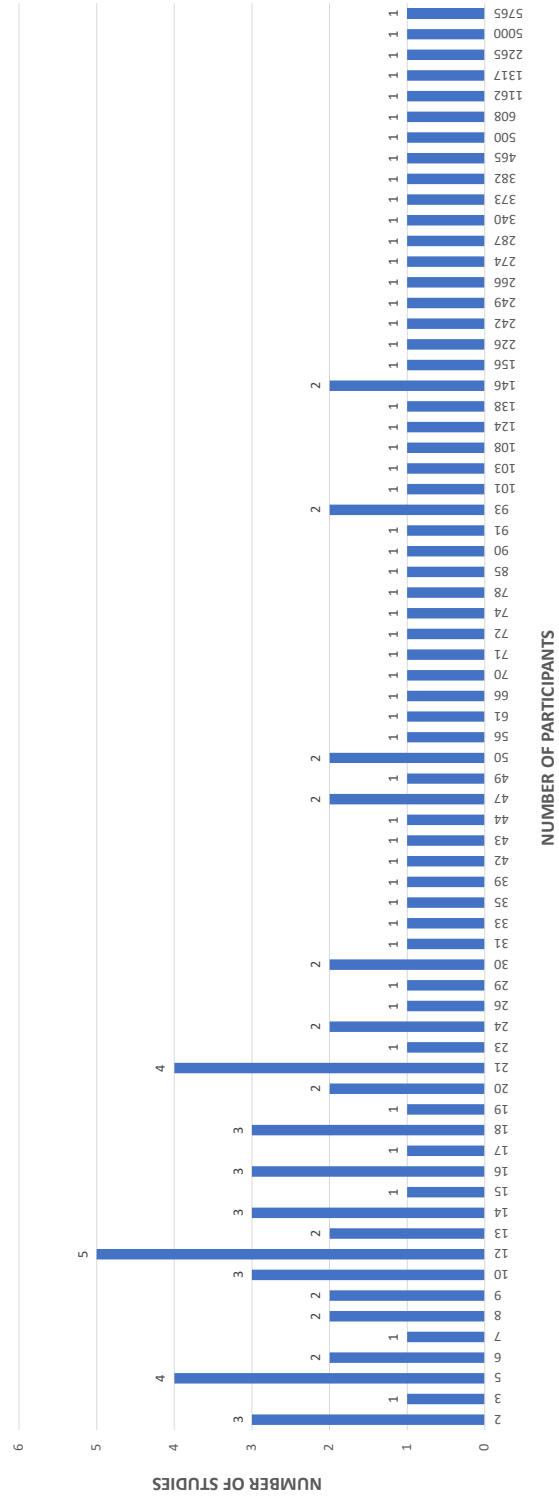
Figure 4.1: Relationship between the number of research participants and the number of studies.

Table 4.1: The frequency in which studies report diversity category (reported column) per type of participant study.

Study Type	# of Studies		# Diversity Categories	
	Total	Reported	Average	Max
Surveys	51	49	3.0	7
Task based	33	32	3.3	8
Interviews	22	22	3.0	7
Validation	19	19	2.8	6
Lab study	9	9	4.6	8
Field study	6	6	2.8	6
Grounded theory	4	4	3.2	6
Control Group	2	2	2.5	3
Action Research	1	1	4.0	4
Focus groups	1	1	2.8	6
Judgement study	1	1	3.0	3

The number of studies surpasses 105, as a participant study may involve multiple study types (e.g., interviews and surveys). The study type is based on how the authors describe their own study.

4.2 RQ2. What diversity categories are reported in SE participant studies?

After determining that most SE participant studies do report aspects of participant diversity, we want to identify what kinds of diversity are reported. Such analysis will help us better understand what characteristics of participant diversity are emphasized by SE researchers and the aspects that might be overlooked and require more attention. To answer RQ2, we created a coding scheme where we identified descriptors about participants in SE studies. We then grouped similar descriptors to form diversity categories as mentioned in section 3.6.

We found 12 different diversity categories reported in SE participant studies. Table 4.2 shows the diversity categories we identify in the content analysis, as well as their definition and frequency (Freq) in our sample. To aid visulization of the frequency of the diveristy categories, we present a graph shown in Figure 4.2. The frequencies do not sum up to 100% as, on average, studies report multiple categories of diversities (as shown in Table 4.1).

Overall, two types of diversity remained dominant in the participant diversity reported by our sample of studies, the *Experience* and the *Main Occupation* of participants. This shows that SE

Table 4.2: Diversity categories identified in 105 studies from ICSE 2019, 2020, and 2021 and their descriptions.

Themes	Diversity Category	Definition	Freq.
Professional	**Experience**	The reported professional working experience of a participant such as their experience in programming languages, industrial work experience, number of years of experience in their field.	89%
	Main Occupation	The current occupation of the participant, ex. developer. The current occupation also includes being in graduate school, for example, if the participant's main occupation is currently pursuing their PhD.	77%
	Education	The reported education of a participant could include their level of education, their educational institution or their field of study.	37%
Social	**Gender/Sex**	The reported gender or sex of participants.	39%
	Location	The reported geographical location of participants.	33%
	Age	The reported age of participants.	18%
	Language	The reported spoken or written language of participants.	8%
	Nationality	The reported nationality of participants.	5%
	Race	The reported race of participants.	4%
	Psychological	The reported psychological measurements and neurological data of participants such as Optimism, Pessimism, Extroversion, Neuroticism, etc.	2%
	Physiological	The reported physiological characteristics such as right-handedness or physical disability such as colour-blindness.	2%
	Socioeconomic status	The reported socioeconomic status of the participant.	1%

studies are often concerned with showcasing a diverse sample of participants regarding professional backgrounds. The *Experience* of participants was reported in 89% of the studies. We group in this category different kinds of professional experiences, from years of industrial experience to experience in a specific technology, such as programming languages. As Zhang, Yang, Lopes, and Kim (2019) report in their study:

> "Eleven participants had **two to five years of Java experience**, while the other five were **novice programmers with one-year Java experience**, showing a good mix of different levels of Java programming experience" (p. 323).

Main Occupation was reported in 77% of the studies, showing a well-established practice to collect and report participation occupations. This category includes professional occupations (e.g., developer, quality analyst) and education-related occupations (e.g., Ph.D. student, MSc. student).

For example, as Ju, Sajnani, Kelly, and Herzig (2021) describes:
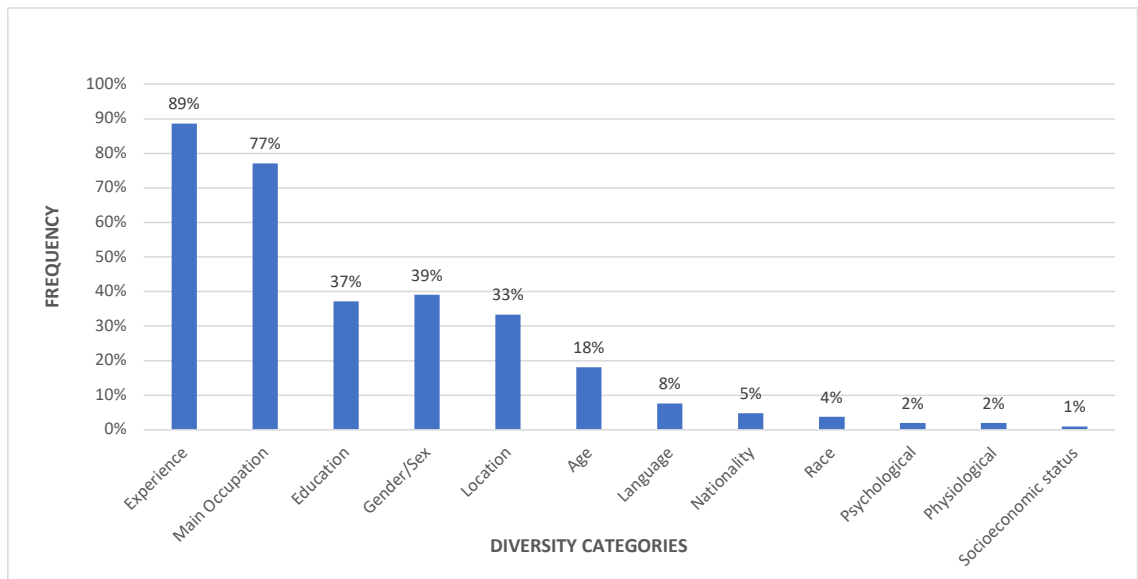
Figure 4.2: The diversity categories identified in participant studies and the frequency at which they were reported in the sample of 105 studies.

> "This process yielded 397 **developers** and 1167 **developer managers** for interviews and 1629 **developers** and 754 **developer managers** for surveys" (p. 614).

Three other categories were reported to describe participants in a comparable frequency: *Gender/Sex* (39%), *Education* (37%), and *Location* (33%). Reported in at least a third of the studies, these categories help qualify the diversity of participants in terms of gender, scholarly level and geographic location. For example, Dias et al. (2021) report the location diversity of their participants as follows:

> "Our interviewees are **located in different countries**, such as Brazil, Canada, Czech Republic, USA, Germany, and Portugal" (p. 983).

Age (18%) and *Language* (8%) of participants are less commonly reported in our sample of SE studies. The age of participants is a characteristic that may be indirectly captured by participants' years of experience. However, it is remarkable that the language spoken by participants is seldom reported, given the international audience of SE papers and the importance of effective communication in any type of study with participants. For example, Xia, Wan, Kochhar, and Lo (2019) reported on the efforts of translating the survey content to ensure effective communication with participants:

> "To support respondents from China, we translated our survey to **Chinese** before publishing the survey" (p. 926).

Finally, some categories were reported only in a handful of SE participant studies. These categories include *Nationality* (5%), *Race* (4%), *Physiological* (2%), *Psychological* (2%) and *Socioeconomic status* (1%). These categories describe specific racial and social aspects of participants and seem to only be collected and reported in specific cases. For example, Krueger et al. (2020) ask participants to complete psychological measurement surveys, such as the Positive and Negative Affect Scale (PANAS, emotional health). Peitek, Apel, Parnin, Brechmann, and Siegmund (2021) reported that all participants in their study had a normal or corrected-to-normal vision and were right-handed.

We identified two major themes of diversity categories: professional and social, among the 12 categories we found (refer to Themes column of Table 4.2).

Professional diversity pertains to a participant's profession or professional experience, such as their occupation, work experience, and education, as represented by the categories *Occupation*, *Experience*, and *Education*. We considered education as professional diversity since we observed that the participants' education was contextualized with their professional experience rather than their social profile.

On the other hand, **social diversity** pertains to demographics and social differences, as represented by the categories *Gender/Sex*, *Age*, *Location*, *Language*, *Nationality*, *Race*, *Socioeconomic status*, *Physiological*, and *Psychological* characteristics.

Our analysis revealed that SE studies tend to prioritize reporting professional diversity categories over social diversity categories when selecting and reporting participants. In fact, all professional categories are reported in at least one-third of all studies, while only *Gender* and *Location* among social categories reach similar levels of frequency. This is significant because social diversity categories, also known as personal or identity categories, are important for equity, diversity, and inclusion efforts (Committee et al., 2021).
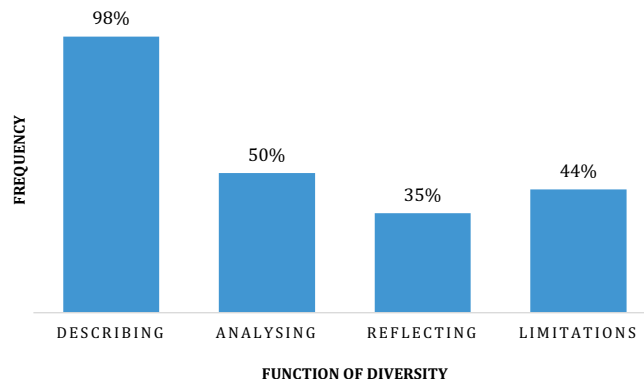
Figure 4.3: How participant diversity is used in the sample of 105 SE studies. We report the frequency of studies that use participant diversity for **describing** participants, **analysing** results, **reflecting** the study outcomes, and assessing the study **limitations**.

4.3 RQ3. What is the function of diversity in SE participant studies?

In our prior research question (RQ), we uncovered a multitude of diversity categories that are frequently reported in SE studies. These studies commonly highlight the professional backgrounds of participants as the most emphasized aspect of diversity.

In this RQ, our focus is to gauge the significance of participant diversity in SE research. To achieve this objective, we have adopted the methodology outlined in Section 3.7. Our methodology classifies the function of participant diversity into four distinct categories:

- Describing: providing a comprehensive description of the participants in the study.

- Analysing: evaluating the impact of diversity on the study outcomes.

- Reflecting: considering the role of participant diversity in the conclusions drawn from the study.

- Limitations: examining the limitations posed by participant diversity in the study.

With this approach, we aim to gain a deeper understanding of the role of participant diversity in SE research and its impact on the validity of the results.

Figure 4.3 shows the frequency of every function of participant diversity in the sample of 105 studies. We determine that 98% of studies describe participant diversity by considering 103 out

Table 4.3: The function of participant diversity in the 105 studies, broken down by diversity category.

Diversity Categories	Describing	Analysing	Reflecting	Limitations
Experience	82%	44%	28%	27%
Occupation	76%	17%	10%	10%
Gender/Sex	37%	10%	6%	10%
Education	35%	8%	5%	8%
Location	29%	9%	4%	13%
Age	17%	4%	2%	2%
Language	5%	1%	0%	4%
Nationality	4%	1%	0%	1%
Race	4%	0%	0%	0%
Psychological	2%	1%	1%	0%
Physiological	1%	1%	1%	0%
Socioeconomic	1%	0%	0%	0%

To better visualize the differences across categories, we highlight the most frequent function per category in dark blue and the second most frequent function in a light blue.

of 105 studies that report diversity. Naturally, this result is expected as the primary function of participant descriptors is to describe the participants. Interestingly, however, participant diversity is less frequently referred to when researchers analyse their study results (50%), reflect upon the results (35%), and assess the limitations of their study (44%). This leads us to conclude that **in most SE participant studies, diversity is reported as a means to describe participants but is less frequently used for further analysis or reflection in the research**.

Table 4.3 displays a breakdown of the participant diversity function, across diversity categories. To make it easier to identify patterns across the diversity categories, we highlight the highest frequency (dark blue) and the second-highest frequency (light blue) in each category. To aid visualization, we present separate graphs of different categories in Figure 4.4, Figure 4.5 and Figure 4.6.

Our findings demonstrate that the category of participant *Experience* is the most consistently used in SE research across all four functions of diversity reporting - description, analysis, reflection, and assessment of limitations. While participant *Occupation* is also a frequently reported diversity category, it is mainly used to simply describe participants, with only a small percentage of studies incorporating it in the analysis or reflection of their results. This can be clearly seen in Figure 4.4.

Figure 4.4: The percent frequency of studies that use the diversity categories experience and occupation for describing participants, analysing results, reflecting upon the study outcomes, and assessing the study limitations. **These two categories present all four functions.**

This highlights the importance that SE researchers place on participant experience, as at least a quarter of the studies in our sample delve into its impact on the results, outcomes, and limitations. On the other hand, all other diversity categories are mainly used for description purposes alone, with little consideration given to their potential impact on the research. Our conclusion is that **SE researchers place significant emphasis on participant experience in their studies, whereas the consideration of other diversity categories remains limited across various functions of diversity**.

Some interesting patterns emerge when we look at the second most frequent function (light blue) per diversity category in Table 4.3. We identified *Gender/Sex*, *Location*, and *Language* as social diversity categories which researchers use more frequently to assess the study limitations, as can be seen in Figure 4.5. Often, these categories are used to discuss the limitations of the study's generalizability (external validity). The following quote from Alsuhaibani, Newman, Decker, Collard, and Maletic (2021) is an example of the use of participation diversity to discuss the study's limitations:

"With regards to external validity, we did not directly collect geolocation data. However, we did collect IP addresses, which gave us country information. From this we determined that participants came from 72 different countries, mainly from Europe and North America. Thus, we feel that the results are generalizable to a broad population" (p. 597).
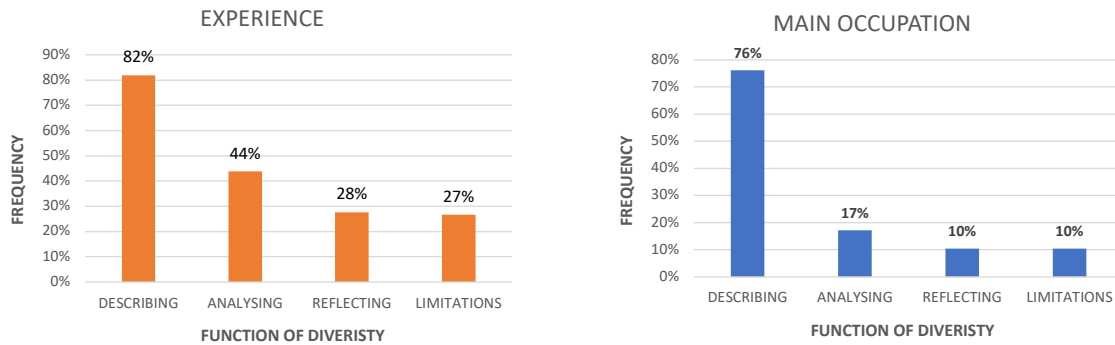
Figure 4.5: The percent frequency of studies that use the diversity categories gender, location, and language for describing participants, analysing results, reflecting upon the study outcomes, and assessing the study limitations. **These three categories are most frequently used to assess limitations.**

It is evident that social diversity categories *Race* and *Socioeconomic status* are underrepresented in SE research. Our analysis revealed that only 4% and 1% of the studies in our sample, respectively, reported on these categories. While these categories were used to describe the participants in these studies, they were not utilized for further analysis or reflection, as can be seen in Figure 4.6.

This finding leads us to the conclusion that although some SE researchers collect and report data about participants' race and socioeconomic status, they do not use this data for further analysis or reflection in their research. This is not necessarily problematic. For example, considering to which extent the sample of research participants represents the demography is not necessarily linked to investigating social differences. However, especially in the case that participants are far from representing the demography, a further reflection on the impact of this fact on the results of the research might be important.

In conclusion, the role of diversity in software engineering (SE) studies can be summarized into four crucial functions: description, analysis, reflection, and assessing limitations. Our findings
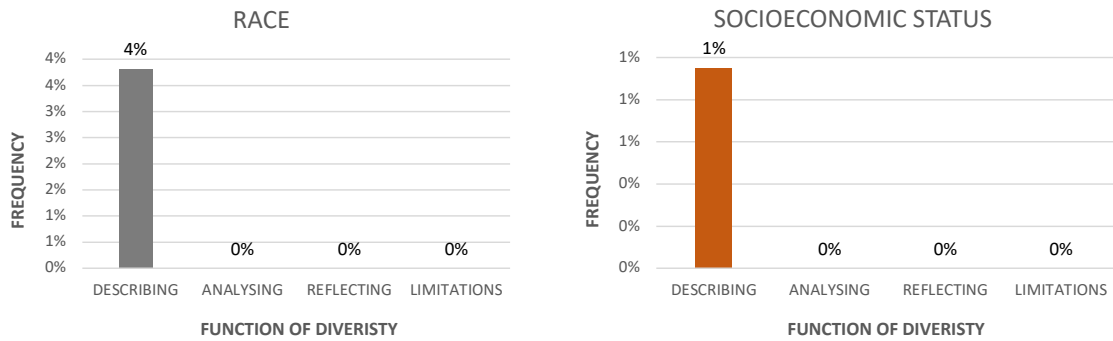
Figure 4.6: The percent frequency of studies that use the diversity categories race and socioeconomic status for describing participants, analysing results, reflecting upon the study outcomes, and assessing the study limitations. **These two categories are used only when describing participants**.

reveal that, although SE researchers generally include information about participant diversity in the description of study participants, they tend to fall short when it comes to integrating diversity considerations into their analysis of study results, reflection on study outcomes, and assessment of the limitations of their studies. This highlights the need for greater attention to diversity in all aspects of SE research, from study design to data interpretation.

Chapter 5

Discussion

In Chapter 5, we introduce diversity awareness models that can assist SE researchers in evaluating their diversity awareness levels. Additionally, we provide guiding questions that can aid SE researchers in integrating diversity considerations into their research. It is important to note here that we are not providing researchers with checklists suggesting the types of diversity categories they should consider in their study. Our focus though this research study is to spread awareness amongst SE researchers to critically consider diversity within their research.

5.1 Diversity Awareness Model

While studying the extent to which SE participant studies report on the diversity of participants, we identified five types of studies which were distinguishable in terms of considering diversity in the overall research and publication approach. Based on these findings, we have proposed a model for diversity awareness, as shown in Figure 5.1. The model encompasses five levels of diversity awareness, providing a useful tool for researchers to assess their own level of diversity awareness and guide their reflections on diversity in their studies.

To illustrate the different levels of diversity awareness, we have provided short fictional examples for each level in Figure 5.1. We refrain from providing real examples for all levels, as our goal is not to single out past studies but to discuss paths for improving future ones.

However, we do believe that good real-world examples can be instructive for SE researchers. As

such, we highlight the work of Dias et al. (2021) as a study with high diversity awareness (level 4). The authors of this study consciously "chose" to prioritize diversity in their participant selection, and they clearly explain their reasoning for doing so as can be seen by the following quote:

> "To foster diversity, when inviting the participants, we prioritize women and non-US based maintainers. We took this decision to avoid having too many "Silicon Valley" participants, as they are over-represented amongst OSS maintainers" (Dias et al., 2021) (p. 983).

It is also worth noting that some studies in our sample went beyond considering only the diversity of research participants, and also reported the diversity of the researchers involved. The study by Gerosa et al. (2021) is a good example of this, as they highlighted the diversity of their research team, which was composed of international researchers from different regions and backgrounds as can be seen from the following quote:

> "we formed an international and diverse team of researchers, who are originally from South America (4), Europe (3), and Asia (1) and were working, at the time of this study, in North America (5), Europe (1), South America (1), and Australia (1). Seven researchers work in academia with extensive experience with OSS, and one researcher is a practitioner working in an OSS company" (Gerosa et al., 2021) (p. 1048).

This demonstrates a level of diversity awareness that extends beyond just participant diversity.

In conclusion, our model for diversity awareness provides a useful framework for SE researchers to assess their own level of diversity awareness and strive for continuous improvement in this area. Through reflective practice, SE researchers can contribute to a more inclusive and equitable software engineering community.

5.2   Diversity Awareness Guidelines

It is important to note that quantifying the level of diversity awareness in research is a complex task. Simply reporting on a few diversity categories or stating that gender diversity was not a consideration in a study does not necessarily equate to a high level of diversity awareness. Instead,

Level 0 (non identifiable diversity awareness)

- No reporting of diversity of participants.
- No reporting of considerations of diversity for the selection of participants.
- No reflection on the lack of diversity (including statements such as "diversity (or gender) was not considered").

Level 1 (beginner's level)

- Mentioning of some diversity categories, categories are limited. ✔
- No further explanation on the relevance of the categories.
- No reporting of considerations of diversity for the selection of participants.
- No reflection on the lack of diversity and implications for research.

Example: "20% of the research participants were women" (no further reflection).

Level 2 (intermediate level)

- Mentioning of diversity categories. ✔
- Reporting of the consideration of diversity for the selection of participants. ✔
- Limited to no reflection on the lack of diversity and implications for research.
- Limited to no explanation on the relevance of the categories.

Example: "Although we tried to reach a balanced representation of gender, only 20% of the research participants were women."

Level 3 (advanced level)

- Reporting the consideration of diversity in the selection of participants. ✔
- Reflecting diversity (or the lack thereof) and implications for research. ✔
- Explaining the relevance of the categories. ✔

Example: "We paid attention to diversity in recruiting participants. Since our study involves reaction time, vision, and other physical factors, we paid special attention to include participants of diverse age and diverse physical ability."

Level 4 (high level)

- Extensive reporting of the consideration of diversity in the selection of participants. ✔
- Extensive reflection on diversity (or the lack thereof), implications for research, improvement of research design. ✔
- Extensive explanation and reflection of the relevance of particular categories for the research topic. ✔

Example: ""We paid attention to diversity in recruiting participants in terms of gender, race, language, age. Since our study involves reaction time, vision, and other physical factors, we paid special attention to include participants of a broad diversity of age and diverse physical ability. However, the study design might have benefited from a greater diversity among the researchers. All researchers involved in this study are between 27-40 years old, 80% of the researchers are male, and no researcher has color vision deficiency or a physical disability."
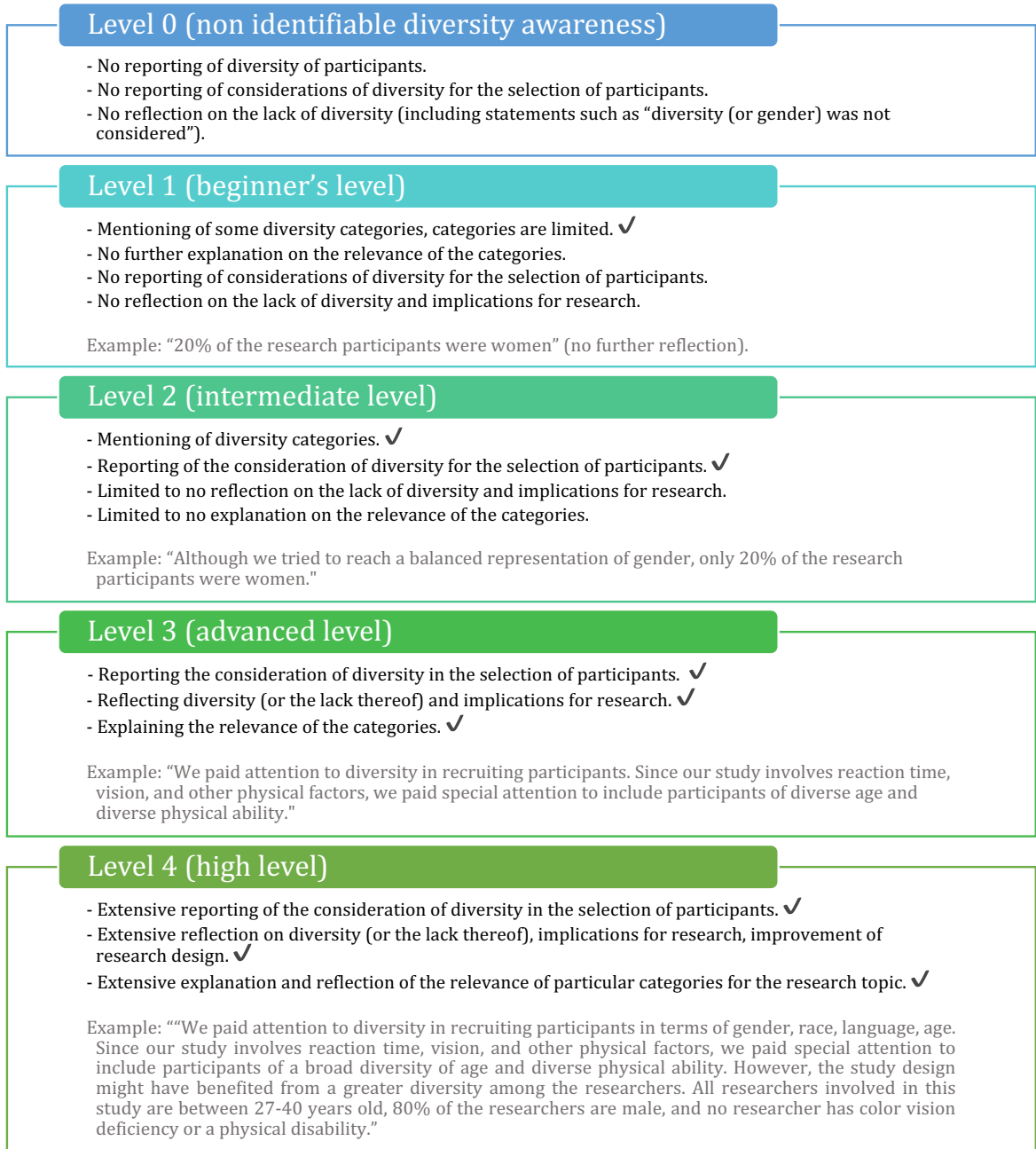
Figure 5.1: **Diversity Awareness Model** for SE participant studies that can be used as a self-assessment tool by SE researchers to guide their reflections on diversity in their studies.

Table 5.1: Diversity Awareness Guidelines for SE researchers.

Guiding Questions for Researchers

- Do you consider diversity in your research?
- Which diversity categories do you consider when selecting research participants?
- Are the diversity categories considered relevant to your research?
- Which diversity categories do you not consider? Why do you not consider them?
- How do you make sure that your research participants are diverse (e.g., a demographic representation of society)?
- What recruitment efforts do you undertake to reach a demographic representation?
- Which groups/diversity categories are over-represented?
- Which professional diversity categories do you consider (e.g., work experience, ...)?
- Which social diversity categories do you consider (e.g., gender, language, nationality, age, ...)?
- Which diversity categories might be relevant in the context of your research?
- Are you reporting your diversity considerations in your research publication?

it is the combination of various actions such as reporting, analysing, addressing, and reflecting on diversity considerations that truly showcases the authors' level of diversity awareness.

To help researchers in their self-assessment of diversity awareness, we have proposed a set of guiding questions in Table 5.1. These questions encourage researchers to reflect on their approach to diversity and consider ways to improve their practices in future studies. Ultimately, it is through continuous reflection, examination and improvement that researchers can increase their diversity awareness and contribute to more inclusive and equitable software engineering research.

Chapter 6

Threats to Validity

In this chapter, we recognize the threats to validity of our research and discuss methods we applied in our research design to mitigate these threats. We discuss three main threats to validity of our results. Firstly, we assess the validity of the results caused by bias or error within the experiment, also known as internal validity. Secondly, we identify the external validity of our results by identifying if our results are generalizable. Finally, we reflect on the construct validity of our results where we assess the limitations of the study design and framework we chose for this study to accurately answer our RQs.

6.1   Internal Validity

The selection process of our sample of participant studies poses a potential threat to the internal validity of our research findings. As described in Section 3.4, I read all the abstracts from the selected venue to identify participant studies for our sample. However, there is a possibility that some participant studies may have been missed due to a lack of methodology description in their abstracts. To mitigate this threat of subjectivity, we took two steps. Firstly, whenever ambiguity arose, I read the entire paper to determine its inclusion in the study. Secondly, in cases where ambiguity persisted, the research team engaged in discussions to arrive at a final decision on the study's inclusion or exclusion. Given our sample consists of 105 participant studies, it is unlikely that any missing study would significantly impact the overall results.

Another potential threat to the internal validity of our findings is the classification of diversity categories in the coding scheme for RQ1 and RQ2. Given that this classification was performed manually, it could be perceived as subjective. To mitigate this risk, we took a number of steps, as outlined in Section 3.5. Firstly, we conducted a pilot study and engaged in regular discussions throughout the study to ensure the consistent classification of each diversity category in our pilot sample. Secondly, in instances of ambiguity during the coding process, we reached consensus through discussions among the authors. These measures helped ensure the validity of our results.

6.2 External Validity

The external validity of our research may be questioned due to the limited scope of our sample selection. Specifically, we chose to focus on the ICSE technical tracks from the years 2019, 2020, and 2021. While ICSE is a highly regarded conference in the software engineering research community and three years of sample papers is a common industry practice (e.g., (Storey et al., 2020)), there is the potential for our results to differ if we had selected a different venue for our sample. This could restrict the generalizability of our findings.

However, despite these limitations, we believe that the core contributions of our work are still valid and applicable. Our creation of a coding scheme for identifying diversity categories, the development of a model for diversity awareness, and the provision of guidelines for diversity awareness represent valuable contributions and have the potential to impact the broader software engineering community.

6.3 Construct Validity

The construct validity of our research is challenged by our coding scheme. We established four distinct functions of diversity - describing, analyzing, reflecting, and assessing limitations - based on our understanding of relevant functions. However, we recognize that alternative characterizations of the functions of diversity in participant studies may exist. Our classification of diversity categories also carries a similar threat, as the categories we reconstructed from our sample could potentially be labeled differently. However, given the absence of existing classification schemes in the field of

software engineering that met our needs, we deemed it necessary to establish our own scheme. We hope that our approach to identifying diversity categories in SE participant studies will serve as a foundation for future works in this area.

Chapter 7

Conclusion, Impact on Society and Future Work

In the concluding chapter, we present a summary of the thesis, reflecting upon its main findings and contribution to the field of SE. Furthermore, we investigate the impact of this research on society, highlighting its noteworthy contributions to addressing critical issues. Lastly, we delve into the prospects for future research and propose potential avenues for further exploration.

7.1 Conclusion

Incorporating diversity considerations in research, development, and innovation has become an increasingly important topic. It is a well-known fact that diverse teams produce better outcomes, whereas the lack of diversity might result in biased and discriminatory technologies. Therefore, the inclusion of diverse stakeholders is considered paramount for the creation of an ethical and socially responsible future. With this study, we aim to contribute to the conversation on how EDI (equity, diversity, inclusion) can be implemented in Software Engineering (SE) research.

In our study, we focus on SE research that includes research participants since this is an evident opportunity to consider diversity, and we investigate to which extent and with what purpose SE researchers consider and report diversity in their research papers. To this end, we examine the

diversity considerations in SE participant studies. We apply content analysis to investigate participant studies from three ICSE technical tracks, from 2019 to 2021. Our investigation focused on understanding 1) the extent to which participants are described, 2) what diversity categories are more prominent in SE research, and 3) the function participant diversity serves in SE studies. Consequently, we examine four main functions of diversity, namely, describing, analysing, reflecting upon and, assessing limitations of the diversity of participants in SE participant research.

Our results demonstrate that only a few studies do not consider diversity at all, however, the examined studies differ greatly in the range of the consideration and reporting of diversity. On one hand, our findings shed light on some positive remarks for the SE research community. Reporting participant diversity is a well-established practice, with studies reporting on multiple characteristics of participants. On the other hand, our results also point to some gaps/challenges that may need further addressing. Studies emphasize participants' professional backgrounds over their social backgrounds, which may prevent important reflections that are needed in a research community.

Furthermore, participant diversity is often only reported initially in the studies, to describe participants, but is seldom analysed or reflected upon when researchers discuss their study results, outcomes and limitations. From these outcomes, we draw the conclusion of differences in the diversity awareness among SE researchers. Finally, we propose a model of diversity awareness for participant studies as a tool to support SE researchers in reflecting on diversity and incorporating it systematically in their research.

7.2   Impact on Society

We believe this research to have a significant impact on the design of future SE research, in particular research that includes research participants. Our work both highlights the importance of considering and reporting the diversity of participants as well as provides guiding questions to successfully integrate considerations of diversity and inclusion into the research design. Furthermore, our proposed model of diversity awareness helps researchers to self-assess and review their research to identify gaps which, otherwise, they would have missed. For research that includes participant studies, defining, selecting and recruiting research participants is an evident opportunity to

consider and implement diversity and inclusion into research. The better diversity is considered and implemented in the research design the broader the acceptance and the benefit of the research and development for society will be.

Additionally, this research helps advance the United Nations sustainable development goals (SDG) (United Nations Department of Economic and Social Affairs, 2015), more specifically:

- **SDG 5 - Gender Equality**, which endeavors to attain gender parity and empower all women and girls. Our work emphasizes the significance of considering diversity in SE research so that underrepresented groups, including women, are not neglected.

- **SDG 9 - Industry, Innovation, and Infrastructure**, which aims to foster inclusive and sustainable industrialization and innovation. Through this research, we aim to ensure that future advancements and innovations in technology and SE consider diversity and inclusion in their design and implementation.

- **SDG 10 - Reduce Inequalities**, which strives towards reducing inequality within and among countries. We want to make sure that products created as a result of SE research are not just created for certain groups of people, such as people from one country, but are created keeping diverse populations in mind.

Overall, our research highlights the importance of considering diversity and inclusion in SE research and provides a framework for researchers to integrate these considerations into their work. By doing so, we aim to promote the development of more ethical, socially responsible, and inclusive technologies that better serve the needs of diverse communities. Our work aligns with the United Nations' sustainable development goals and has the potential to make a meaningful contribution to a more equitable and sustainable world.

7.3   Future Work

Our study is a crucial step forward in the conversation about diversity and inclusion in SE participant studies. It sheds light on the existing strengths and limitations of diversity awareness in

these studies and provides a model that future studies can leverage to enhance their level of diversity awareness.

The potential for further exploration in this area is substantial. Future research could assess the impact of neglecting diversity considerations in SE participant research. By leveraging our research methodology, combining both qualitative and quantitative analysis, it would be possible to gain new insights into the prevalent diversity categories across various types of SE research. This study can also serve as a blueprint for similar research to be conducted in other engineering fields, offering insights into the state of diversity and inclusion in those fields.

It is our belief that the Diversity Awareness Model presented in our study has the potential to be applied to a variety of other engineering fields and drive positive change towards more equitable and inclusive research practices.

References

Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). Sampling: Why and how of it. *Indian Journal of Medical Specialties*, *4*(2), 330–333.

Akrich, M. (1992). *The de-scription of technical objects.* MIT press.

Albusays, K., Bjorn, P., Dabbish, L., Ford, D., Murphy-Hill, E., Serebrenik, A., & Storey, M.-A. (2021). The diversity crisis in software development. *IEEE Software*, *38*(2), 19-25. doi: 10.1109/MS.2020.3045817

Al-Nashif, N. (2021). *A human rights-based approach to equitable access to stem education* (1st, New ed. ed.; T. Tajmel, K. Starl, & S. Spintig, Eds.). Münster/New York: Waxmann.

Alsuhaibani, R. S., Newman, C. D., Decker, M. J., Collard, M. L., & Maletic, J. I. (2021). On the naming of methods: A survey of professional developers [Book Section]. In *Proceedings of the 43rd international conference on software engineering* (p. 587–599). IEEE Press. Retrieved from https://doi.org/10.1109/ICSE43902.2021.00061 doi: 10.1109/icse43902.2021.00061

Assembly, U. G., et al. (1948). Universal declaration of human rights. *UN General Assembly*.

Baltes, S., & Ralph, P. (2022). Sampling in software engineering research: A critical review and guidelines. *Empirical Software Engineering*, *27*(4), 1–31.

Blackstone, A. (2018). *Principles of sociological inquiry: Qualitative and quantitative methods*. Saylor Academy Open Textbooks.

Blaisdell, S. (1994). Factors in the underrepresentation of women in science and engineering: A review of the literature. *Women in Engineering ProActive Network*.

Bosu, A., & Sultana, K. Z. (2019, Sep.). Diversity and inclusion in open source software (oss)

projects: Where do we stand? In *2019 acm/ieee international symposium on empirical software engineering and measurement (esem)* (p. 1-11). doi: 10.1109/ESEM.2019.8870179

Buolamwini, J., & Gebru, T. (2018, Feb). Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Proceedings of the 1st conference on fairness, accountability and transparency* (Vol. 81, p. 77-91). PMLR. Retrieved from https://proceedings.mlr.press/v81/buolamwini18a.html

Burnett, M., Peters, A., Hill, C., & Elarief, N. (2016). Finding gender-inclusiveness software issues with gendermag: A field investigation. In *Proceedings of the 2016 chi conference on human factors in computing systems* (p. 2586–2598). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2858036.2858274 doi: 10.1145/2858036.2858274

Catolino, G., Palomba, F., Tamburri, D. A., Serebrenik, A., & Ferrucci, F. (2019). Gender diversity and women in software teams: How do they affect community smells? In *Proceedings of the 41st international conference on software engineering: Software engineering in society* (p. 11–20). IEEE Press. Retrieved from https://doi.org/10.1109/ICSE-SEIS .2019.00010 doi: 10.1109/ICSE-SEIS.2019.00010

Collins, P. H., & Bilge, S. (2020). *Intersectionality*. John Wiley & Sons.

Committee, C. R. C., et al. (2021). Best practices in equity, diversity and inclusion in research. *Government of Canada*, *22*. Retrieved from https://www.sshrc-crsh.gc.ca/ funding-financement/nfrf-fnfr/edi-eng.aspx

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. , *1*, 139–167.

Damian, D., & Zeller, A. (2022). Message from the icse 2022 program chairs. In *2022 ieee/acm 44th international conference on software engineering (icse)* (p. xxxi-xxxii).

Dias, E., Meirelles, P., Castor, F., Steinmacher, I., Wiese, I., & Pinto, G. (2021). What makes a great maintainer of open source projects? In *2021 ieee/acm 43rd international conference on software engineering (icse)* (pp. 982–994).

Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., Kirn, A., . . . Potvin, G. (2016). More comprehensive and inclusive approaches to demographic data collection.

Fricker, R. D. (2008). Sampling methods for web and e-mail surveys. *The SAGE handbook of online research methods. London: SAGE Publications Ltd*.

Garvie, C., Bedoya, A., & Frankle, J. (2019). *The perpetual line-up. unregulated police face recognition in america. georgetown law center on privacy & technology.*

Gerosa, M., Wiese, I., Trinkenreich, B., Link, G., Robles, G., Treude, C., . . . Sarma, A. (2021). The shifting sands of motivation: Revisiting what drives contributors in open source. In *Proceedings of the 43rd international conference on software engineering* (p. 1046–1058). IEEE Press. Retrieved from https://doi.org/10.1109/ICSE43902.2021.00098 doi: 10.1109/ICSE43902.2021.00098

Hankivsky, O. (2014). Intersectionality 101, 1–34. *Canada: The Institute for Intersectionality Research & Policy, SFU*.

Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.

Hearn, J., & Louvrier, J. (2015). Theories of difference, diversity, and intersectionality. *The Oxford handbook of diversity in organizations*, 62–82.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? women in science, technology, engineering, and mathematics.* ERIC.

Icse '20: Proceedings of the acm/ieee 42nd international conference on software engineering. (2020). New York, NY, USA: Association for Computing Machinery.

Ju, A., Sajnani, H., Kelly, S., & Herzig, K. (2021). A case study of onboarding in software teams: Tasks and strategies. In *2021 ieee/acm 43rd international conference on software engineering (icse)* (pp. 613–623).

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* [Book]. SAGE Publications. Retrieved from https://books.google.ca/books?id=nE1aDwAAQBAJ

Krippendorff, K., & Bock, M. (2009). *The content analysis reader* [Book].

Krueger, R., Huang, Y., Liu, X., Santander, T., Weimer, W., & Leach, K. (2020). Neurological divide: An fmri study of prose and code writing. In *Proceedings of the acm/ieee 42nd*

international conference on software engineering (p. 678–690). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3377811.3380348 doi: 10.1145/3377811.3380348

Lambrecht, A., & Tucker, C. E. (2018). Algorithmic bias? an empirical study into apparent gender-based discrimination in the display of stem career ads. *An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads (March 9, 2018)*.

Lenarduzzi, V., Dieste, O., Fucci, D., & Vegas, S. (2021). Towards a methodology for participant selection in software engineering experiments: A vision of the future. In *Proceedings of the 15th acm / ieee international symposium on empirical software engineering and measurement (esem).* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3475716.3484273 doi: 10.1145/3475716.3484273

Litvin, D. R. (1997). The discourse of diversity: From biology to management. *Organization*, *4*(2), 187–209.

McIlwee, J. S., & Robinson, J. G. (1992). *Women in engineering: Gender, power, and workplace culture*. SUNY Press.

McKinzie, A. E., & Richards, P. L. (2019). An argument for context-driven intersectionality. *Sociology Compass*, *13*(4), e12671.

Menezes, A., & Prikladnicki, R. (2018). Diversity in software engineering. In *Proceedings of the 11th international workshop on cooperative and human aspects of software engineering* (p. 45–48). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3195836.3195857 doi: 10.1145/3195836.3195857

Miller, C., Rodeghero, P., Storey, M.-A., Ford, D., & Zimmermann, T. (2021). How was your weekend? software development teams working from home during covid-19. In *Proceedings of the 43rd international conference on software engineering* (p. 624–636). IEEE Press. Retrieved from https://doi.org/10.1109/ICSE43902.2021.00064

Nagappan, M., Zimmermann, T., & Bird, C. (2013). Diversity in software engineering research. In *Proceedings of the 2013 9th joint meeting on foundations of software engineering* (pp. 466–476).

Nkomo, S., & Stewart, M. (2006, 01). Diverse identities in organizations. In (p. 520-540).

NSERC. (2022). *Nserc guide on integrating equity, diversity and inclusion considerations in research.* Retrieved 2022-09-22, from https://www.nserc-crsng.gc.ca/NSERC -CRSNG/Policies-Politiques/EDI_guidance-Conseils_EDI_eng.asp

Oberleitner, G. (2021). *Understanding the human right to science: Cescr general comment no. 25 (2020)* (1st, New ed. ed.; T. Tajmel, K. Starl, & S. Spintig, Eds.). Münster/New York: Waxmann.

Oudshoorn, N., Rommes, E., & Stienstra, M. (2004). Configuring the user as everybody: Gender and design cultures in information and communication technologies. *Science, Technology, & Human Values*, *29*(1), 30–63.

Padala, H. S., Mendez, C., Fronchetti, F., Steinmacher, I., Steine-Hanson, Z., Hilderbrand, C., ... Sarma, A. (2022, jan). How gender-biased tools shape newcomer experiences in oss projects. *IEEE Trans. Softw. Eng.*, *48*(1), 241–259. Retrieved from https://doi.org/ 10.1109/TSE.2020.2984173 doi: 10.1109/TSE.2020.2984173

PAR-L. (2010). *Calls to action: Human rights complaint concerning the canada research chairs program. policy, action, research list.* Retrieved 2022-09-22, from https://www2.unb .ca/parl/CRC_index.htm

Parliament, E., & Council. (2013). Regulation (eu) no 1291/2013 of the european parliament and of the council of 11 december 2013 establishing horizon 2020-the framework programme for research and innovation (2014-2020) and repealing decision no 1982/2006/ec. *Official Journal of the European Union*, *20*(12).

Patrick, H. A., & Kumar, V. R. (2012). Managing workplace diversity: Issues and challenges. *Sage Open*, *2*(2), 2158244012444615.

Pauly Morgan, K. (1996). Describing the emperor's new clothes: three myths of educational (in) equality. *Diller A et al. The gender question in education: theory, pedagogy and politics. Boulder. CO: Westview*, *272*.

Peitek, N., Apel, S., Parnin, C., Brechmann, A., & Siegmund, J. (2021). Program comprehension and code complexity metrics: An fmri study. In *Proceedings of the 43rd international conference on software engineering* (p. 524–536). IEEE Press. Retrieved from https://doi .org/10.1109/ICSE43902.2021.00056 doi: 10.1109/ICSE43902.2021.00056

Perera, H., Hussain, W., Whittle, J., Nurwidyantoro, A., Mougouei, D., Shams, R. A., & Oliver, G. (2020). A study on the prevalence of human values in software engineering publications, 2015 – 2018. In *Proceedings of the acm/ieee 42nd international conference on software engineering* (p. 409–420). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3377811.3380393 doi: 10.1145/3377811.3380393

Pieterse, V., Kourie, D. G., & Sonnekus, I. P. (2006). Software engineering team diversity and performance. In *Proceedings of the 2006 annual research conference of the south african institute of computer scientists and information technologists on it research in developing countries* (p. 180–186). ZAF: South African Institute for Computer Scientists and Information Technologists. Retrieved from https://doi.org/10.1145/1216262.1216282 doi: 10.1145/1216262.1216282

Rodriguez, G., Nadri, R., & Nagappan, M. (2021, 09). Perceived diversity in software engineering: a systematic literature review. *Empirical Software Engineering*, *26*. doi: 10.1007/s10664-021-09992-2

Saldaña, J. (2021). The coding manual for qualitative researchers. *The coding manual for qualitative researchers*, 1–440.

Sarewitz, D. (2005, 01). This won't hurt a bit: Assessing and governing rapidly advancing technologies in a democracy. *The Future of Technology Assessment*, 14-21.

Schelenz, L. (2022). Diversity concepts in computer science and technology development: A critique. *Science, Technology, & Human Values*, 01622439221122549.

Schiebinger, L., Klinge, I., Paik, H. Y., Sánchez de Madariaga, I., Schraudner, M., & Stefanick, M. (2011-2020). *Gendered innovations in science, health medicine, engineering, and environment.* Retrieved 2022-09-22, from http://genderedinnovations.stanford.edu/index.html

Shaw, M. (2003). Writing good software engineering research papers: Minitutorial. In *Proceedings of the 25th international conference on software engineering* (p. 726–736). USA: IEEE Computer Society.

Storey, M. A., Ernst, N. A., Williams, C., & Kalliamvakou, E. (2020, sep). The who, what, how of

software engineering research: A socio-technical framework. *Empirical Softw. Engg.*, *25*(5), 4097–4129. Retrieved from https://doi.org/10.1007/s10664-020-09858-z doi: 10.1007/s10664-020-09858-z

Tamtik, M. (2022, Feb.). Mirroring society? tracing the logic of diversity in the canadian journal of higher education. *Canadian Journal of Higher Education*, *51*(3), 127–144. Retrieved from https://journals.sfu.ca/cjhe/index.php/cjhe/article/view/189125 doi: 10.47678/cjhe.vi0.189125

Tannenbaum, C., Ellis, R., Eyssel, F., Zou, J., & Schiebinger, L. (2019, 11). Sex and gender analysis improves science and engineering. *Nature*, *575*, 137-146. doi: 10.1038/s41586-019-1657-6

United Nations Department of Economic and Social Affairs. (2015). *Sustainable development goals.* Retrieved from https://sdgs.un.org/goals

United Nations Statistics Division. (n.d.). *Population Censuses' Datasets (1995 - Present).* Accessed Feb, 25, 2023 [Online]. Retrieved from https://unstats.un.org/unsd/demographic-social/products/dyb/index.cshtml/#censusdatasets

Van Oudheusden, M., & Shelley-Egan, C. (2021, 11). Rri futures: learning from a diversity of voices and visions. *Journal of Responsible Innovation*, *8*, 1-9. doi: 10.1080/23299460.2021.1989656

Vertovec, S. (2012). "diversity" and the social imaginary. *European Journal of Sociology / Archives Européennes de Sociologie*, *53*(3), 287–312. doi: 10.1017/S000397561200015X

Wu, H., Deng, W., Niu, X., & Nie, C. (2021). Identifying key features from app user reviews. In *Proceedings of the 43rd international conference on software engineering* (p. 922–932). IEEE Press. Retrieved from https://doi.org/10.1109/ICSE43902.2021.00088 doi: 10.1109/ICSE43902.2021.00088

Xia, X., Wan, Z., Kochhar, P. S., & Lo, D. (2019). How practitioners perceive coding proficiency. In *2019 ieee/acm 41st international conference on software engineering (icse)* (pp. 924–935).

Zanoni, P., & Janssens, M. (2004). Deconstructing difference: The rhetoric of human resource managers' diversity discourses. *Organization studies*, *25*(1), 55–74.

Zhang, T., Yang, D., Lopes, C., & Kim, M. (2019). Analyzing and supporting adaptation of online code examples. In *2019 ieee/acm 41st international conference on software engineering (icse)*

(pp. 316–327).

Zhao, D., Xing, Z., Chen, C., Xia, X., & Li, G. (2019). Actionnet: Vision-based work-flow action recognition from programming screencasts. In *Proceedings of the 41st international conference on software engineering* (p. 350–361). IEEE Press. Retrieved from https://doi.org/10.1109/ICSE.2019.00049 doi: 10.1109/ICSE.2019.00049