



# PUC

ISSN 0103-9741

Monografias em Ciência da Computação

nº 03/2021

## **On Psychometric Instruments in Software Engineering Research: An Ongoing Study**

**Danilo Almeida Felipe**

**Marcos Kalinowski**

Departamento de Informática

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO**

**RUA MARQUÊS DE SÃO VICENTE, 225 - CEP 22451-900**

**RIO DE JANEIRO - BRASIL**

# On Psychometric Instruments in Software Engineering Research: An Ongoing Study

Danilo Almeida Felipe, Marcos Kalinowski

dfelipe@inf.puc-rio.br, kalinowski@inf.puc-rio.br

**Abstract.** [Context] Although software development is an inherently human activity, research in software engineering (SE) has long focused mostly on processes and tools, failing to recall about the human factors behind. Even when explored, researchers typically do not properly use psychology background to better understand human factors in SE, such as the psychometric instruments, which aim to measure human factors. [Objective] Our goal is to provide a critical review on the use of psychometric instruments in SE research regarding personality. [Method] We present a two-step study. First, a systematic mapping of the literature in order to generate a catalog of the psychometric instruments used; second, a preliminary survey to be conducted with social sciences researchers to assess their adoption in SE research. [Results and Conclusion] The results so far are quite initial. The next steps direct us to finish the data extraction to finalize the catalog (systematic mapping) and to refine the survey design and apply it with social sciences researchers.

**Keywords:** behavioral software engineering; personality; mapping study; survey

**Resumo.** [Contexto] Embora o desenvolvimento de software seja uma atividade humana, a pesquisa em Engenharia de Software (ES) tem concentrado esforços majoritariamente em processos e ferramentas, esquecendo-se dos fatores humanos por trás. Ainda quando explorados, os pesquisadores não tem adotado adequadamente referencial da psicologia para entender melhor os fatores humanos em ES, bem como dos instrumentos psicométricos, que visam medir algum tipo de fator humano. [Objetivo] Nosso objetivo é fornecer uma revisão crítica sobre o uso de instrumentos psicométricos nas pesquisas em ES em relação à personalidade. [Método] Apresentamos um estudo em duas etapas. Primeiro, um mapeamento sistemático da literatura para gerar um catálogo dos instrumentos psicométricos utilizados; segundo, um survey preliminar a ser realizado com pesquisadores de ciências sociais/humanas para avaliar sua adoção na pesquisa em ES. [Resultados e Conclusões] Os resultados até agora são iniciais. Os próximos passos nos direcionam a finalizar a extração de dados para finalizar o catálogo (mapeamento sistemático) e refinar o survey para conduzi-lo com pesquisadores de ciências sociais/humanas.

**Palavras-chave:** engenharia de software comportamental; personalidade; mapeamento da literatura; survey

**In charge of publications:**

PUC-Rio Departamento de Informática - Publicações

Rua Marquês de São Vicente, 225 - Gávea

22453-900 Rio de Janeiro RJ Brasil

Tel. +55 21 3527-1516 Fax: +55 21 3527-1530

E-mail: [publicar@inf.puc-rio.br](mailto:publicar@inf.puc-rio.br)

Web site: <http://bib-di.inf.puc-rio.br/techreports/>

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background and Related Work</b>	<b>2</b>
2.1	Behavioral Software Engineering (BSE) . . . . .	2
2.2	Psychometrics in Software Engineering . . . . .	3
<b>3</b>	<b>Systematic Mapping Protocol</b>	<b>3</b>
3.1	Mapping Goal and Research Questions . . . . .	4
3.2	Search Strategy and Data Source . . . . .	4
3.2.1	Existing Mapping Study and the Need to Update . . . . .	4
3.2.2	Strategy to Collect New Evidence . . . . .	6
3.3	Study Selection . . . . .	7
3.4	Data Extraction and Classification Scheme . . . . .	7
3.5	Preliminary Results and Next Steps . . . . .	8
3.5.1	Answers to the RQs . . . . .	9
<b>4</b>	<b>Preliminary Survey Protocol</b>	<b>16</b>
4.1	Population and Sample . . . . .	17
4.2	Instrument Design . . . . .	17
4.3	Data Analysis . . . . .	18
4.4	Current Status and Next Steps . . . . .	18
<b>5</b>	<b>Conclusions and Next Steps of The Study</b>	<b>20</b>
	<b>References</b>	<b>22</b>
	<b>Annex A List with the final set of included studies (ongoing)</b>	<b>25</b>

# 1 Introduction

Software Engineering (SE) activities are primarily performed by humans. However, many empirical studies have only focused on proposing new methods and technologies to support SE activities leaving human and social factors behind them underexplored [1], not allowing a more holistic view of the area.

Behavioral Software Engineering (BSE) is the body of knowledge of SE research that attempts to understand human aspects related to the activities of software engineers, software developers, and other stakeholders [2]. The topic has been the subject of recent research in the SE domain. Nevertheless, because it is relatively immature, some approaches adopted misled researchers mainly by not properly combining SE research with social sciences backgrounds, such as psychology, to address human factors [3]. Graziotin *et al.* [4] enlighten a set of existing research on relating developer happiness to productivity, software quality, and social interactions.

Furthermore, in SE research, it is inherent to note that measurement activities are an essential part. In empirical studies the researcher(s) must be sure when adopting ways of measuring the study’s resources in question. These resources can be personnel, hardware, or software for an activity or process [5]

BSE research has encouraged the use of psychometric instruments as support to the understanding of human factors in a more systematic way [1, 2]. In its turn, Psychoempirical Software Engineering proposed in Graziotin *et al.* [3] deals with “denoting research in Software Engineering with proper theory and measurements from psychology”. A problem addressed by the authors is the misuse of theoretical backgrounds of psychology, such as assuming a theory as the only truth in the research foundation; and also the improper use of psychometric instruments, some not validated from psychology or used to evaluate wrong human factors. The authors propose guidelines to assist researchers in addressing these problems.

However, information about how psychometric instruments are adopted in SE research remains vague and dispersed in many studies. As far as we know, one study has partially synthesized this knowledge concerning a specific construct, personality<sup>1</sup>: a period of forty years (1970 - 2010) about personality in SE research is mapped by Cruz *et al.* [6]. Still, there is only a brief discussion and characterization of the instruments and their use in the software engineering context (*i.e.*, education and pair programming), missing a critical assessment. This status quo remains unchanged more than ten years later and deserves to be challenged.

There is a need to find out whether SE research over the years has been adopting these instruments coherently with their proposed form of use in the social sciences. Although some authors have outlined this on a smaller scale [7, 8], a large-scale study has not been done to get a big picture.

In order to synthesize this knowledge in a structured manner, **our objective is to present an overview on the use of psychometric instruments in SE research on personality**<sup>2</sup>, with a focus on classifying the *type of research, empirical evaluations, objective of the studies, and purposes of the psychometric instruments*. We intend to consolidate our findings in a catalog through a systematic mapping study (secondary study). Also,

---

<sup>1</sup>Personality is one of the most studied concepts in BSE research as pointed by Lenberg *et al.* [2].

<sup>2</sup>Hereafter we commonly refer to psychometric instruments related to personality in SE research simply as “psychometric instrument(s)”.

to assess the use of these instruments from the point of view of social sciences researchers by conducting a survey (primary study).

More specifically, we address the following Research Questions (RQ):

- **RQ1:** *What are the psychometric instruments used regarding personality in SE research?*
- **RQ2:** *How do social sciences researchers perceive the adoption of psychometric instruments regarding personality in SE research?*

To the best of our knowledge, this study is the first one focused on creating a catalog of psychometric instruments used in SE research and assessing them with social sciences researchers in a larger scale. The remainder of this study manuscript is organized as follows. Section 2 presents the background concepts and related work. Section 3 presents the systematic mapping protocol and preliminary results. Section 4 presents our preliminary survey protocol. Finally, section 5 presents our concluding remarks.

## 2 Background and Related Work

This section provides background information and related work.

### 2.1 Behavioral Software Engineering (BSE)

BSE is defined as “the study of cognitive, behavioral and social aspects of software engineering performed by individuals, groups or organizations” [2]. It involves dealing with existing relationships between SE and disciplines from social sciences, such as *work and organizational psychology*, the *psychology of programming*, and *behavioral economics* to get a broader understanding of SE practices.

Although software is developed by humans, for a long-time, SE research has focused intensively on the technical aspects (such as processes and tools) and less on human and social aspects [1]. A summarization conducted in Graziotin *et al.* [9] points out existing research, in the scope of BSE, on relating developer happiness to productivity, software quality, and also social interactions.

Still, in the context of BSE, Lenberg *et al.* [2] present a definition for the body of knowledge research described earlier and also conducted a systematic literature review based on the definition. The findings report lack of research in some SE knowledge areas (*i.e.*, requirements, design, and maintenance) and rare collaboration between SE and social science researchers.

Cruz *et al.* [6] performed a systematic mapping on personality in SE research. In addition to reporting on the most common SE topics addressed, such as *education* and *extreme programming*, they report which personality tests were most commonly used, which resembles this work. However, the authors only collected and reported brief information on personality-related psychometric instruments without deep discussion about them regarding their use in SE. They provide a valuable list of instruments and relate them to some SE topics, but lack a critical review involving the lens of social scientists.

Further, Usman and Minhas [8] investigate ethical topics raised by McDonald and Edwards [7] on the adoption of MBTI (Myers-Briggs Type Indicator) based tests, a psychometric instrument, in a sample of 8 studies obtained in the final set of Cruz *et al.* [6]

published after 2007, and complemented with 7 studies returned in string-based search in Scopus<sup>3</sup> in the years of 2016 and 2017, totaling a sample of 15 studies. The authors found problems in all of them, including the reliability and validity of MBTI (there are different versions of this psychometric instrument). The authors also highlight possible causes, such as not exploring literature guidelines and lack of collaboration with social science researchers. However, the study reported is initial and limited to analyzing only the use of MBTI in a small sample of studies.

Our study intends to conduct an extensive analysis, with more cataloged psychometric instruments, exploring their relation with SE theory and using expert knowledge to assess these instruments' use.

## 2.2 Psychometrics in Software Engineering

According to Michell [10], “psychometrics is concerned with theory and techniques for quantitative measurement in psychology and social sciences” [Michell, 1999 *apud* Feldt et al., 2008], in addition Feldt *et al.* [1] state that “... in practice, this often means the measurement of knowledge, abilities, attitudes, emotions, personality, and motivation”. In the present study, we focused on personality.

The use of psychometric instruments in SE has been encouraged, especially in empirical research, as a way to emphasize hitherto unexplored human factors and to help understand how they affect the research landscape [1].

As far as we know, McDonalds and Edwards [7] is the first study that manifests attention on the use of psychometric instruments in SE research, beyond providing guidelines for the use of two of them (MBTI and 16PF). In addition, one of the authors is from social sciences and a certified professional regarding these instruments.

Still under discussion, Graziotin *et al.* [3] claim that the use of psychometric instruments should be cautious, recommending that they should be validated; in addition to the proper theoretical background used. The authors then propose the *Psychoempirical Software Engineering* that aims “to denote research in SE with proper theory and measurement from psychology”. In the same study, the authors provide guidelines for the proposal.

More recently, Usman and Minhas [8] show results, despite several limitations, that the use of psychometric instruments in SE is inadequate. The authors emphasize who should apply them, but we consider that trained professionals are the best decision according to the little evidence and guidelines in literature so far.

Although we are researchers without formal qualifications in social sciences, we rely on consolidated methodological tools of SE to conduct this study and plan to involve social science scientists in our research. Given the gathering of evidence collected to answer **RQ1**, we intend to use the experts' opinion to answer **RQ2** in order to allow for a critical assessment. The next section presents our systematic mapping protocol and preliminary results.

## 3 Systematic Mapping Protocol

Systematic mapping is a method to build a classification scheme of an area providing a visual summary of the state of research in a structured way [11]. It aims at providing an

---

<sup>3</sup><https://www.scopus.com/>

auditable and replicable process with minimal bias.

This section describes each step of our research method based on guidelines in the literature. Subsection 3.1 introduces the mapping goal and research questions. Subsection 3.2 describes the search strategy for collecting new evidence. Subsection 3.3 presents the study selection criteria and discusses quality assessment. The subsection 3.4 presents the Data Extraction Form and our classification scheme, and subsection 3.5 presents preliminary results.

### 3.1 Mapping Goal and Research Questions

Our systematic mapping aims at providing an overview of the use of psychometric instruments in SE research. To guide our investigation, and to obtain an overview of the state-of-the-art, trends and gaps, we describe the main Research Question (RQ) as follows:

**RQ1:** *What are the psychometric instruments used regarding personality in SE research?*

Based on the main RQ, we derived five secondary RQs in order to further characterize the field as follows.

- **RQ1a:** *What are the objectives of the studies?*
- **RQ1b:** *For what purposes have psychometric instruments been used in the studies?*
- **RQ1c:** *Which types of research do the studies refer to?*
- **RQ1d:** *Which types of empirical studies have been conducted?*
- **RQ1e:** *What are the limitations faced by the use of psychometric instruments reported in the studies?*

In the following subsection the search strategy is presented. It was developed by the first author and reviewed by the second author.

### 3.2 Search Strategy and Data Source

#### 3.2.1 Existing Mapping Study and the Need to Update

This mapping started being conducted in a standard way of conducting secondary studies (string-based search in digital libraries with snowballing steps), according to consolidated literature [11, 12, 13, 14]. Until new guidelines emerged, and we noticed that they could help conduct our study [15, 16], given the awareness we had about comprehensive secondary studies on human factors in SE that could be updated [6, 2].

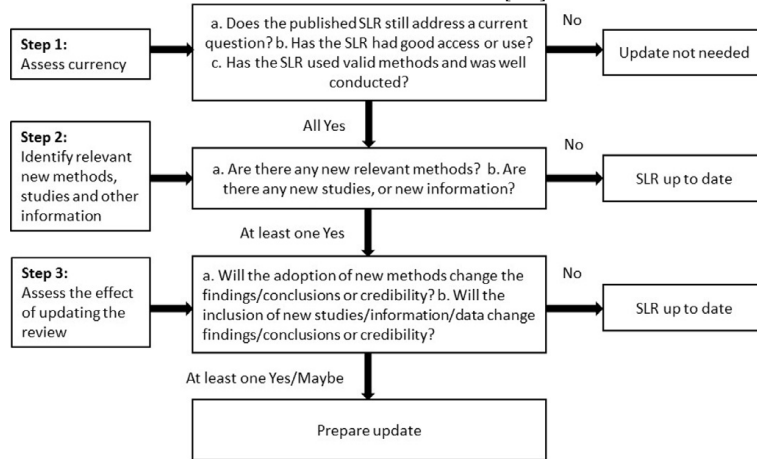
We defined Cruz *et al.* [6] as a candidate for updating as we wanted to start our immersion into BSE using a narrower scope, focused on personality, to allow a complete overview and a precise critical assessment. Cruz *et al.* [6] identified 90 papers published within a time range of 40 years, more than 10 years ago. Lenberg *et al.* [2] define 55 BSE concepts (*e.g.*, *personality*, *job satisfaction*, *communication*, etc.) in a large scale study that considered 250 papers. The narrower focus by Cruz *et al.* [6] would also allow us to apply our updating search strategy (discussed in subsection 3.2.2) with reasonable efforts.



We also argue that personality is a BSE concept presented in [2] as one of the most studied together with others (such as *group composition*, *communication*, and *organizational culture*). We believe that updating Cruz *et al.* [6] yields significant results regarding our objectives described in section 1, which is also within BSE’s scope. Thus, we decided to update the mapping study by Cruz *et al.* [6].

To evaluate the need of updating Cruz *et al.* [6], we used the framework recommended by guidelines to evaluate the possibility of updating secondary studies in SE [15]<sup>4</sup>. We conducted the evaluation process by answering the same seven RQs used in Mendes *et al.* [15], listed and answered hereafter (Steps 1.a. to 3.b.) and illustrated in Figure 1.

Figure 1: Framework recommended by Mendes *et al.* [15], adapted from Garner *et al.* [17]



**Step 1.a.:** *Does the published SLR still address a current question?*

Yes. Human factors, such as personality, have been the subject of SE research despite being historically poorly explored [4]. A list of arguments follows: recent efforts to consolidate a body of knowledge in SE as proposed by Lenberg *et al.* [2] defining BSE; existing conferences such as CHASE (Workshop on Cooperative and Human Aspects of Software Engineering), subsidized by ICSE (International Conference on Software Engineering)<sup>5</sup>, the largest conference in SE; award-winning papers on the topic in main software engineering related venues, such as [4], awarded in the Journal of Systems and Software<sup>6</sup>. All these arguments show the importance of human factors as a relevant research topic in SE.

**Step 1.b.:** *Has the SLR had good access or use?*

Yes. Like Mendes *et al.* [15], we used in the cut-off point the same yearly average citation value of 6.82 documented by Garousi and Fernandes [18] to consider a paper for good access or use. In August of 2020, Cruz *et al.* [6] had a yearly average citation value of 30.8 in Google Scholar.

**Step 1.c.:** *Has the SLR used valid methods and was it well-conducted?*

Yes. Regarding the methods, Cruz *et al.* [6] present an extension of preliminary results published in Cruz *et al.* [19] with some improvements, such as a refined search string increasing the sensitivity and coverage; adding backward snowballing steps; review of RQs

<sup>4</sup>This study proposes guidelines for updating Systematic Literature Reviews (SLR). Despite this, we believe that our mapping goal is comprehensive enough to apply in the context of systematic mappings.

<sup>5</sup><http://www.icse-conferences.org/>

<sup>6</sup><https://www.journals.elsevier.com/journal-of-systems-and-software>

and extended presentation of results. Finally, the authors present clear steps in their mapping protocol and are based on well-recognized guidelines for secondary studies in SE [13].

**Step 2.a.:** *Are there any new relevant methods?*

Yes. Concerning new methods about our mapping protocol, we adopted guidelines presented as the best way to search for evidence to update secondary studies in SE [15]. We believe that we used good literature references to help answer our RQs and, consequently, reflect on the results' presentation. However, different from Cruz *et al.* [6] mapping study, our focus is totally on the psychometric instruments within SE research, not a characterization of an aspect (personality) in general.

**Step 2.b.:** *Are there any new studies or new information?*

Yes. The papers included in the original study had each a considerable number of citations in a preliminary verification in Google Scholar. In addition to having a five-year time interval since the publication of the mapping and the conduction of the update in this present study (2015 to 2020), beyond the period of ten years (2011 to 2020), not incorporated by Cruz *et al.* [6].

**Step 3.a.:** *Will the adoption of new methods change the findings, conclusions or credibility?*

Yes, potentially. We adopted a new method concerning the mapping protocol and addressed different RQs to get a big picture of psychometric instruments in SE research, which we believe generates new and important findings. Cruz *et al.* [6] has a relevant RQ on the psychometric instruments/personality tests used, but there is little discussion of its results beyond listing and counting frequencies.

**Step 3.b.:** *Will the inclusion of new studies/information/data change findings, conclusions or credibility?*

Yes. Regarding new potential findings, we had prior knowledge of a series of studies [3, 9, 4] used as control papers. These studies are not covered by Cruz *et al.* [6] because they were published later. They discuss the use of psychometric instruments in SE, and on theoretical basis of other areas (such as social sciences and psychology), which can support SE research in general.

Next, we describe the search strategy to collect new evidence from a secondary study update in SE.

### 3.2.2 Strategy to Collect New Evidence

We adopted the guidelines proposed in Wohlin *et al.* [16] as a strategy to search for new evidence from a secondary study update. They are the following:

- **Use a seed set containing the original secondary study and its included primary studies:** Cruz *et al.* [6] included 90 papers in their final set. However, one of them was excluded (S86) because it is a book chapter, and we did not find evidence of publication in a scientific journal or conference to be approved in IC1 (see Table 1). As suggested, the secondary study itself was included, obtaining a seed set of 90 studies.
- **Use Google Scholar to search for papers and apply Forward Snowballing (FS), without iteration:** We use the Publish or Perish 7 tool [20] to assist this

step. The tool has features related to bibliometric analysis, in which one is to retrieve citations from publications using Google Scholar (a FS feature). Thus, we conducted the FS in the seed set using the tool in August 2020 and exported the results for treatment in JabRef, a bibliographic reference manager. Also, a new FS step was conducted in January 2021 to ensure full 2020 indexation. All screening steps were conducted using JabRef.

- **Include more than one researcher in the initial screening to minimize the risk of removing studies that should be included (false negatives):** One researcher was included to assist in the initial screening of studies and discussions were held with a third researcher.

### 3.3 Study Selection

Petersen *et al.* [12] argue that only studies that are relevant to answer the RQs must be considered. The inclusion and exclusion criteria applied to filter the raw set of studies from FS are presented in Table 1.

Table 1: Inclusion and exclusion criteria.

Criteria	Description
IC1	Papers published in journals and conferences reporting software engineering research using psychometric instruments regarding personality that were published after 2010.
EC1	Papers that are not written in English.
EC2	Grey literature. Such as books, theses (bachelor’s degree, MSc or PhD), technical reports, occasional papers, and manuscripts without peer-review evidence.
EC3	Papers that are only available in the form of abstracts, posters, and presentations.
EC4	Papers that did not include in their title or abstract terms defined in Cruz <i>et al.</i> [6] regarding personality. There are: “personality”, “psychological typology”, “psychological types”, “temperament type”, and “traits”.
EC5	Studies addressing other psychometric constructs ( <i>e.g.</i> , behavior, cognition, abilities, roles, etc.)

The purpose of this mapping study is to provide an overview of the use of psychometric instruments in SE research published in peer-reviewed venues. Therefore, we focus on classifying the type of contribution, the use of psychometric instruments, and the type of research to understand the overall publication landscape, without applying a formal quality assessment. The procedure here involves reading titles and abstracts and looking for evidence of psychometric instruments. If it is not enough for clarification, the paper’s introduction and conclusion will be read. Still, if not sufficient, the full text of the study will be read.

### 3.4 Data Extraction and Classification Scheme

The data extracted from each paper of the final set is shown in Table 2.

Table 2: Data Extraction Form

Information	Description
Study Metadata	Paper title, author’s information, venue, psychometric instrument (name, version, and application process), and year of publication.
Objective (RQ1a)	Study objective: we will use open coding [21] to extract information.
Purpose of the psychometric instrument in the study (RQ1b)	What constructs represent the purpose of the psychometric instrument in the study. In SE, constructs are derived from one of the classes: <i>people</i> , <i>organizations</i> , <i>technologies</i> , <i>activities</i> , or <i>software systems</i> [22]. We will use open coding [21] to extract data.
Research Type (RQ1c)	For research type facets we use the taxonomy proposed by Wieringa <i>et al.</i> [23], containing the following categories: <i>evaluation research</i> , <i>solution proposal</i> , <i>validation research</i> , <i>philosophical paper</i> , <i>opinion paper</i> , or <i>experience paper</i> . Petersen <i>et al.</i> [12] guidelines are used to assist in this categorization.
Empirical Evaluation (RQ1d)	Classification of the empirical study in the following categories of Wohlin <i>et al.</i> [5]: <i>experiment/quasi-experiment</i> , <i>case study</i> , or <i>survey</i> .
Limitations (RQ1e)	Limitations on the use of psychometric instruments (if exists). Such as what were the difficulties of adoption/application and data interpretation. We will use open coding [21] to extract data.

### 3.5 Preliminary Results and Next Steps

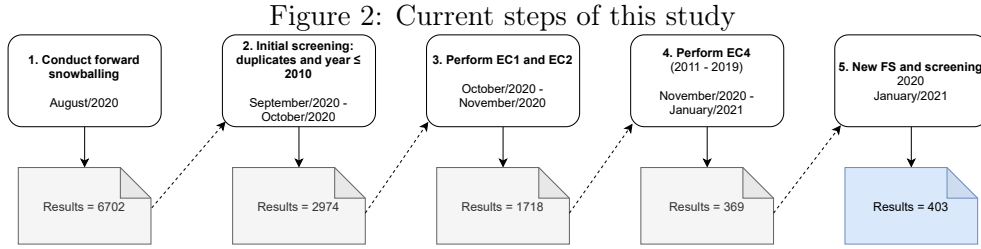
The first step was to conduct FS in the seed set as described in section 3.2.2, which generated an entire of 6702 entries (step 1 of Figure 2). Between September and October of 2020, the first author conducted an initial screening of duplicate and year less than or equal to 2010, given that Cruz *et al.* [6] cover a range from 1970 to 2010.

Many entries were provided by Google Scholar/Publish or Perish 7 export feature with incomplete or incorrect data (*e.g.*, journal studies categorized in the entry as books or miscellaneous, or truncated title or abstracts). After removing duplicate entries and when possible, the first author manually collected the data for publication study and registered it in JabRef to perform a more reliable exclusion per year. The result of this initial screening resulted in 2974 entries (step 2 of Figure 2).

Thereafter, the first author conducted another screening regarding the exclusion criteria EC1 and EC2. The removal was performed based on the metadata provided in the title, abstract, and journal/booktitle field entries. When it was not possible to easily identify, a verification was made through the URL of the entry or searching the source on the web. This exclusion was conducted between October and November of 2020 (step 3 of Figure 2). Each entry was analyzed individually; the exclusion steps mentioned here reduced the set by 1718 entries.

To conduct the EC4 completely, we deleted the existing 2020 entries (which resulted in 369 entries of 2011 to 2019, step 4 of Figure 2) and replaced them with the new 2020

FS ones to get a full ten-year index coverage. We applied the same previous ECs in this step, which resulted in a candidate set of 403 entries (step 5 of Figure 2).



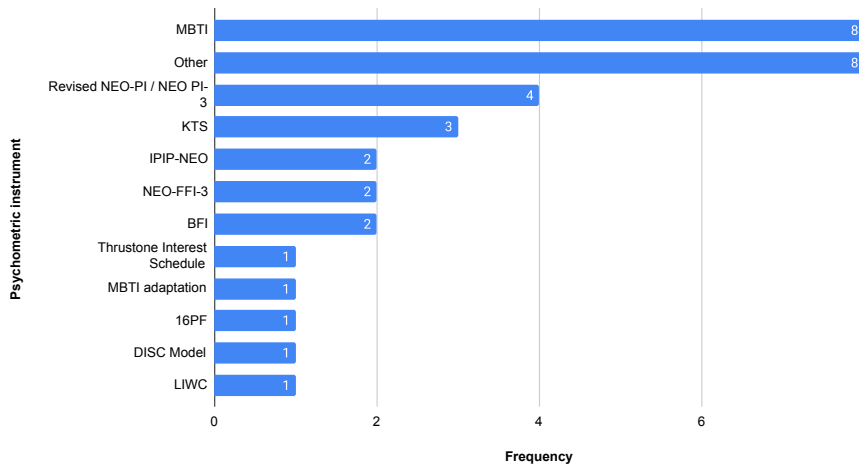
In the following we present some preliminary results extracted and organized by our RQs. So far, we mapped the years from 2011 to 2013<sup>7</sup>. All the steps mentioned above involved the guidance and agreement of the second author. References of the selected studies are presented in Annex A.

### 3.5.1 Answers to the RQs

**RQ1:** *What are the psychometric instruments used regarding personality in SE research?*

We have identified 34 studies that employ psychometric instruments or personality theories, ranging from 2011 to 2013. The frequency of instruments is illustrated in Figure 3, the most used ones were different versions of the Myers-Briggs Type Indicator (MBTI) [S91, S92, S96, S97, S98, S107, S110, S120], as can also be seen in Cruz *et al.* [6].

Figure 3: Frequencies of psychometric instruments



In sequence *Other*, theories of personality such as the Five-Factor Model (also know as Big Five) are employed: [S94] uses it as a basis for the coding process in questionnaire responses; [S104] maps relationships between theory dimensions with soft and hard skills required from software engineers; [S116] is similar to the previous one, but the study uses

<sup>7</sup>We refer to the identification of studies from our protocol as S(NUMBER), where (NUMBER) begins to account from 91 given that Cruz *et al.* [6] has 90 studies.

subjects' social media posts metadata as a source. *Other* was also used to categorize an instrument for which it was not possible to identify a version, bibliographic references, or obtain access [S102, S106, S109, S117] or that discusses the misuse in using psychometric instruments [S119].

To provide an overview of the guidelines raised by McDonalds and Edwards [7], we have also extracted information from the details of types of instruments used (actual name, version, and theory) and how they were employed (from administration to data interpretation). Most studies did not report any bibliographic references and explicit versions of the instrument.

Furthermore, most studies also did not report anything different from “we use x to measure personality” regarding instruments application. Unlike them, [S93] mentions a consent form to participate in the experiment; [S95, S101] report that participants were given instructions; [S114, S118] report that interpretation procedures followed recommendations of the instruments' version. Beyond, [S114] claims that the results were given to the participants. [S124] reports that interviews were conducted following the administration to obtain results but without qualification details of the interviewer.

**RQ1a:** *What are the objectives of the studies?*

During the data extraction, it was possible to observe three major objectives.

*Characterize software engineer personality:* some studies aimed to discover the personality of the software engineer [S91, S96, S97], and to compare them with the personality of professionals from other disciplines [S94, S95]; to discover personality patterns through a methodology [S98, S99]; to associate personality traits to roles in the software development activities (*e.g.*, developer, requirements analyst, tester) or required skills [S99, S104, S107, S116, S117, S123, S124]. Cultural differences were also investigated [S120].

*Predicting team performance:* four studies used information on personality traits (along with other information or not) to predict a development team's performance to optimize resource use. One case reports for academic purposes, combining academic info to build teams to carry out a project [S92]. Others propose tools and methods using machine learning [S103], computational intelligence [S105], and rule-based approaches [S106].

*Investigate the effect of personality:* data on personality traits are used as an intervention to investigate phenomena. We identified this objective with several specificities. Studies, such as [S93, S101, S112, S115], investigate the performance of compositions of pair programming teams; [S100] the influence on the quality of software developed and satisfaction of work perceived by a team; [S102] the programmer's error-proneness while developing; [S108, S110] the use of support tools; [S111] academic team achievements; [S113, S122] predict preferences on development method or mode of working (either face-to-face or virtual); [S114] the performance of software testing; [S118] team members on success in a software project; [S121] implementing new technology. Also, there are discussions concerning future Automated Personality Classification research [S109].

**RQ1b:** *For what purposes have psychometric instruments been used in the studies?*

In Tables 3 to 6 we summarize the constructs to which the psychometric instruments are related by archetypal classes. We highlight the use of the framework to describe theories [22] and its archetypal classes as support for open coding [21] constructs: an *actor* applies *technology* (we believe that *intervention* is more appropriate in the context of psychometric

instruments) to perform certain *activities* on a *software system*.

Figure 4 depicts the relationship network between the constructs described in Tables 3 to 6. It is possible to observe that constructs of class **actor** *researcher* and *academic setting* are the majority, indicating the application of studies in an academic setting or for purely investigative purposes by researchers, which corroborates the classification seen in RQ1c and RQ1d. In the **intervention** class, most data from personality traits measured by some psychometric instrument are used for investigation, followed by its increase with other data (such as leadership style, cognitive style, and skills required) indicated by the construct *personality traits data\**. Regarding the **activity** class, the constructs *software engineering characterization* and *team building* are largely related to **interventions** of *personality traits data*, holding the highest frequency of studies.

Still, these **activities** described earlier are also strongly related to the constructs of the **software system** class. It is possible to observe that *software engineer characterization* is typically not related to any specific software system (*i.e.*, *None* in Figure 4), indicating no direct reference to software system in these studies. The software system code *Tool* indicates the use of some technique using software/games/logic rules to supported the studies. Moreover, *class assignments* were specially related to *team building* and *team building*, where teams were built based in academic contexts based on personality data.

Please note that RQ1b aims at answering what parts of SE theory the psychometric instruments are related to. There may be similarities with the overall objectives of the conducted studies (RQ1a), but RQ1b is specifically focused on the instruments and their relations to SE theory elements.

Figure 4: SE constructs related to the psychometric instruments

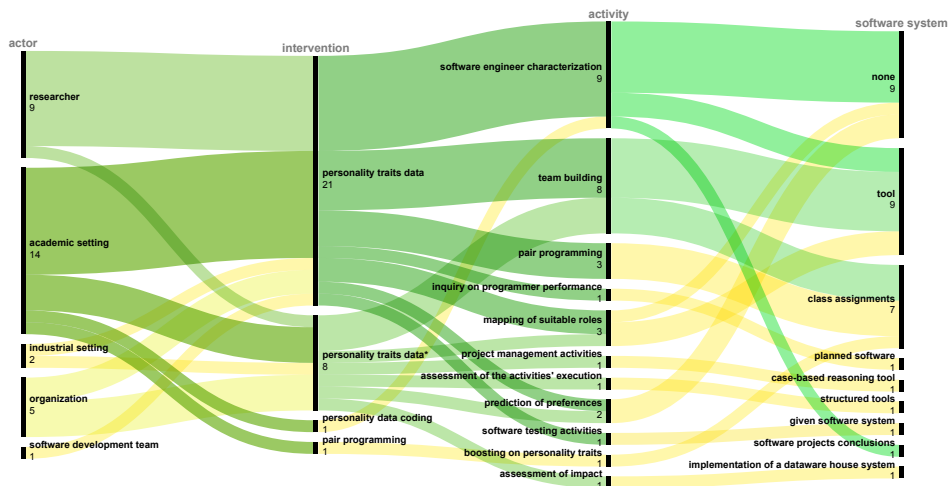


Table 3: Actor

Class	Description	Count
Researcher	The study author(s) act primarily for investigative purposes [S91, S96, S97, S104, S113, S117, S118, S120, S123].	9
Academic setting	The study has an educational purpose or is applied in an academic setting due to scope limitations [S92, S93, S94, S100, S101, S107, S110, S111, S112, S114, S115, S116, S122, S124].	14
Industrial setting	The study is applied in a industrial setting [S95, S121].	2
Organization	The scope of the study is not clear (where does the study data come from?) [S98, S102, S103, S105, S108]. The code was adopted to be comprehensive.	5
Software development team (SDT)	An SDT was the interventor in the study [S99].	1

Table 4: Intervention

Class	Description	Count
Personality traits data	The study has data on personality traits measured by some psychometric instrument [S91, S93, S95, S96, S97, S98, S99, S100, S101, S102, S107, S111, S112, S113, S114, S117, S118, S120, S123, S124] or other source [S116].	21
Personality traits, academic, and software development methodology data	The study has as main input data of personality traits measured by some psychometric instrument, among other data [S92].	1
Personality data coding	Personality data were used by means of coding based on some personality theory [S94].	1
Personality traits and skills data	The study has as main input data of personality traits measured by some psychometric instrument and skills required for some software engineer role [S103, S104].	2
Personality traits and nature of tasks data	The study has as main input data of personality traits measured by some instrument and already mapped information of personality traits recommended for execute a task in a software development process [S105].	1
Personality traits and comprehension performance data	The study has as main input data of personality traits measured by some psychometric instrument and comprehension performance data [S110]	1
Personality traits and leadership style data	The study has as main input data of personality traits measured by some psychometric instrument, among other data [S121]	1
Personality and cognitive traits data	The study has as main input data of personality and cognitive traits measured by some psychometric instrument [S108, S122].	2
Pair programming	The study adopted pair programming to assess some impact [S115].	1



Table 5: Activities

<b>Class</b>	<b>Description</b>	<b>Count</b>
Software engineer characterization	The study characterizes, in some extent, the software engineering professional. It means: comparison of SE and other professionals [S94, S95], and discover (and/or compare) personalities [S91, S96, S97, S98, S118, S120, S123].	9
Team building	The main activity of the study is to build software development teams with more than two members [S92, S99, S100, S103, S105, S107, S111, S124].	8
Pair programming	The main activity of the study is to build pair programming teams [S93, S101, S112].	3
Inquiry on programmer performance	The main activity of the study is to inquiry programmer performance [S102].	1
Mapping of suitable roles	The main activity of the study is to mapping software engineer roles (developer, tester, project manager, etc.) [S104, S116, S117].	3
Project management activities	The main activity of the study is to assess project manager performance given a software system [S108].	1
Assessment of the activities' execution	The main activity of the study is to assess software artifacts and process in a software activity [S110].	1
Prediction of preferences	The main activity of the study is to predict some preference of software engineer professional in a software development activity [S113, S122].	2
Software testing activities	The main activity of the study is to assess tester performance given a software [S114].	1
Boosting on personality traits	The main activity of the study is to boost personality traits at some level [S115].	1
Assessment of impact	The main activity of the study is to assess impacts [S121].	1

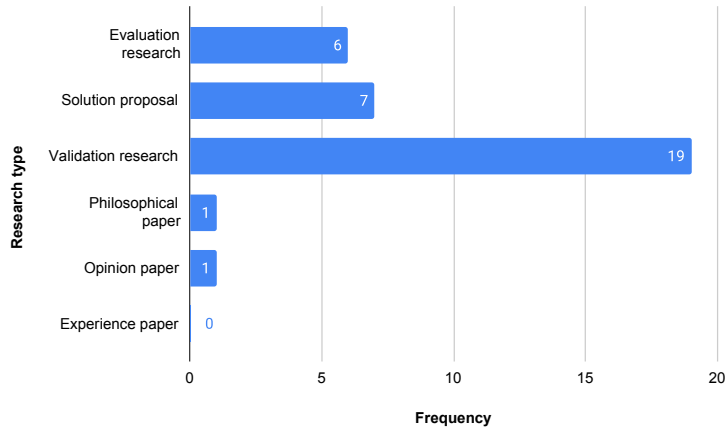
Table 6: Software system

<b>Class</b>	<b>Description</b>	<b>Count</b>
None	No software system was used in the study [S91, S94, S95, S96, S97, S104, S113, S120, S122].	9
Tool	The study used a machine learning/computational intelligence [S92, S98, S103, S105] some kind of algorithm [S116, S117, S123, S124], or a gamecard [S99].	9
Class assignments	The study used a software system for academic settings, previously developed for some specific purpose (be tested, refactored, ...) or developed during the conduction of the study by students [S93, S100, S101, S107, S111, S112, S115].	7
Planned software	The study uses software (or requirements of it) that has not yet been developed, but it was during the conduction of the study [S102].	1
Case-based reasoning (CBR) tool	The study used a CBR tool to perform an investigation [S108].	1
Structured tools	The study used structured tools to perform an investigation [S110].	1
Given software system	A specific software system was used in the study because it has some kind of problem to be investigated by the subjects. For instance: in software testing [S114].	1
Software projects conclusion	The study used data from conclusions of software system projects (successfully or not) [S118].	1
Implementation of a dataware house system	The study used data from implementation of a dataware house system [S121].	1

**RQ1c:** Which type of research do the studies refer for?

Figure 3 shows the distribution of research types facets. It is possible to observe that *validation research* (19 out of 32) overrepresented the set of mapped studies. This facet includes empirical studies, as the less frequent *evaluation research* facet (6 out of 32). The difference indicates that most empirical research has been conducted in academic scenarios for initial validation purposes and does not evaluate something in industrial scenarios. Some *solution proposals* have also been mapped (7 out of 32), which typically represent new proposals with some limited evaluation required for publication.

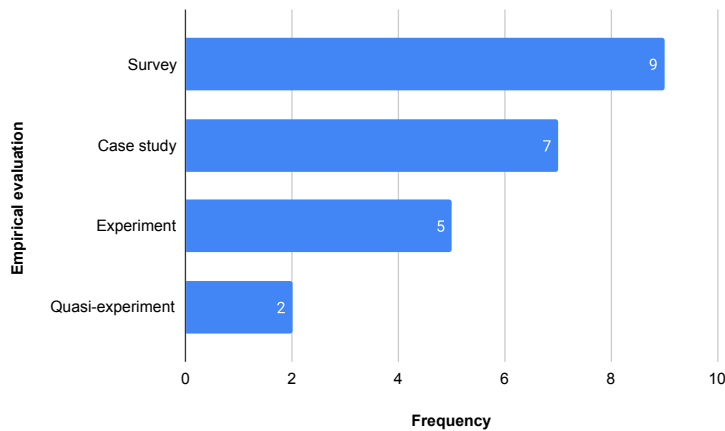
Figure 5: Frequency of research type facets of Wieringa *et al.* [23]



**RQ1d:** Which types of empirical studies have been conducted?

Figure 6 depicts the frequency of empirical evaluations adopted (23 out of 32 studies). It is possible to observe that most of the studies (9 out of 23) used surveys, mostly to apply the psychometric instruments. Still, it is not possible to draw consistent conclusions, such as trends and gaps in empirical evaluation adoption given that we only mapped three years so far.

Figure 6: Frequency of empirical evaluations adopted [5]



**RQ1e:** *What are the limitations faced by the use of psychometric instruments reported in the studies?*

Some studies (16 out of 32) reported limitations related to adoption of psychometric instruments. An overview of these limitations is described below.

*Possible misuse of psychometric instrument:* the authors of [S100] declare that adopting the psychometric instrument to build teams can affect the construct validity, but this limitation is mitigated by relying on the literature. Another study that we consider noteworthy to mention reports on the wide and mistaken use of MBTI for professional guidance in careers in computing [S119]. The authors report a discussion with experts that it could be an “abuse of the MBTI licensing requirements” and ineffective for the purpose.

*Bias in subject responses:* the authors indicate that subjects’ self-administration of psychometric instruments can become a threat if there is no honesty in their responses [S101, S112, S122]. This limitation is mitigated in these studies by ensuring they made aware that the response data obtained is anonymized and used only for research purposes.

*Statistical power of psychometric instrument:* Regarding this limitation, the increase of personality traits data by adopting more psychometric instruments for better prediction is reported [S98, S124]. Other authors also report on replacing the instrument in future work, given that it was not possible to draw significant conclusions in the use of a specific instrument [S113].

Still, regarding the statistical structure of the instrument, [S108] uses an instrument based on the Five-Factor Model and recommends in future works to use the MBTI. In turn, [S112] uses an instrument based on Jung’s theory (KTS / MBTI), which provides strict personality measurements in dichotomies (*e.g.*, introvert-extrovert). This leads to more than one person being categorized in one dichotomy even if they hold different levels. Hereafter, we can notice inconsistencies in the literature regarding the statistical power of the instruments reported in the limitations of the studies.

*No representative data input/no representative sample:* Restrictions of representative data to predict team performance are reported [S92, S105]; the sample size was sufficient to control more than one personality dimension in the experiment, only the Openness to Experience dimension of the Five-Factor Model was considered [S93]; in [S111, S113], a larger sample could add more statistical power to provide other conclusions; limitations in the empirical method employed did not allow representative statistical analysis to generalize results [S118, S122].

*Paid subjects:* Participants were paid to participate in the study, which may have influenced them somehow [S114].

## 4 Preliminary Survey Protocol

This section presents our preliminary survey protocol (primary study) that we aim to conduct to achieve the main objective of this study described in the introduction. We adopted guidelines for conducting surveys in SE [24] and advice to overcome common challenges [25], to assist in developing this protocol. This protocol was developed by the first author and reviewed by the second.

The survey’s main goal is to collect the point of view of social science researchers on the use of psychometric instruments in SE, as mapped through our ongoing secondary study described in section 3. We adopted the Goal-Question-Metric (GQM) method [5] to

characterize the goal in Table 7:

Table 7: Goal of the survey

<b>Analyze</b>	the adoption of psychometric instruments regarding personality in SE
<b>For the purpose of</b>	characterization
<b>With respect to</b>	their correct use
<b>From the point of view of the</b>	social sciences researchers
<b>In the context of</b>	the mapped SE research

Hence, given that there is a poor collaboration between SE and social sciences researchers, and misuse of psychometric instruments, as pointed out in sections 1, 2.1, and 2.2, we have elaborated the following RQ in order to address the survey’s main goal:

**RQ2:** *How do social sciences researchers perceive the adoption of psychometric instruments regarding personality in SE research?*

The remainder of this section is organized as follows. Subsection 4.1 presents information on population and sample. Subsection 4.2 presents the preliminary survey instrument design. Subsections 4.3 and 4.4 discuss data analysis and current status, respectively.

#### 4.1 Population and Sample

The population selected for the study are social science researchers with an area of interest in psychometrics. Initially, we are planning to collaborate with social science researchers from the Department of Psychology at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio)<sup>8</sup> to help us to best plan the outline the population and identify a representative sample.

A priori, the sample may be characterized as being non-probabilistic and accidental, as frequent in SE, where researchers recruit subjects based on personal connections [24]. To characterize a sample frame as ideal, we use points highlighted by Linaker *et al.* [24].

- **Target audience:** social science researchers;
- **Unit of analysis:** psychometric instruments adopted in SE research regarding personality;
- **Unit of observation:** social sciences researchers with an area of interest in psychometrics;
- **Search unit:** to be defined;
- **Source of sampling:** to be defined.

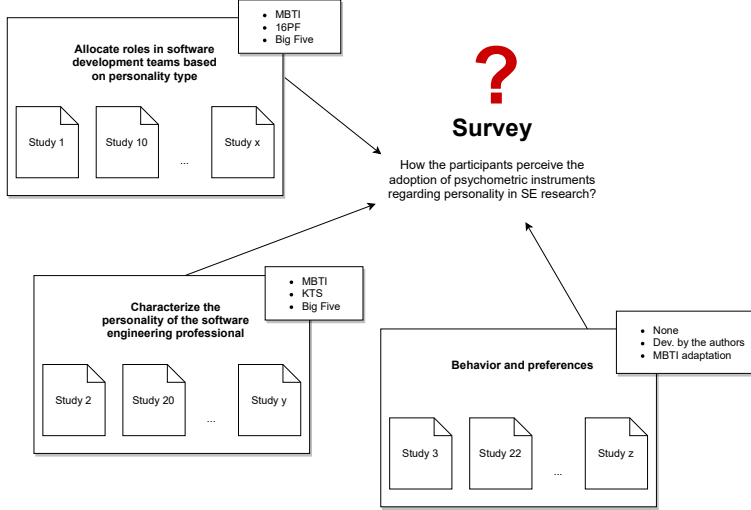
#### 4.2 Instrument Design

The input to conduct the survey consists of completing the systematic mapping. Thus, to assist in surveying the subjects, we intend to cluster the final set of primary studies by the psychometric instrument’s purposes identified through the answered systematic mapping

<sup>8</sup><http://www.psi.puc-rio.br/>

**RQ1.** Figure 7 illustrates this: with each identified instrument (MBTI, 16PF, and a Big Five based instrument), and a set of identified purposes (*e.g.*, *allocate roles in software development teams based on personality type*), and the list of bibliographic references (if the subject needs to consult details of a specific study).

Figure 7: Illustration of clustering of the systematic mapping study artifacts



We rely on the Technology Acceptance Model (TAM) [26] to assess the psychometric instruments and their use in SE research, as we see them as technologies/tools. Thus, the survey questionnaire presented in Table 8 aims to capture the *perception of their usefulness* (questions *Q5* and *Q6*), and *perception of ease of use* (questions *Q7* and *Q8*). Questions *Q1-Q4* concern demographic information on the social sciences researcher (subject) that we intend to collect. Finally, questions *Q9* and *Q10* capture additional limitations regarding the application process and recommendations.

### 4.3 Data Analysis

As presented in Table 8, in our preliminary questionnaire, we only deal with rational data with regard to the experience in years. For such data, we intend to use the mean and standard deviation. All the remaining questions concern ordinal scales (*e.g.*, academic degree, or Likert scales) or open-ended questions. We intend to analyze the agreement of the subjects, in the ordinal data, by counting frequencies, using the median (central tendency) and mode (most frequent answer), potentially also using the median absolute deviation (depending on the sample size). We do not discard using Bootstrapping confidence intervals, but this will heavily depend on our sample size. Cohen's Kappa will be applied for inter-rater consistency evaluations. As for the data visualization, we will use histograms and pie charts to analyze the distribution of answers. For open-ended questions, we will use open (and potentially axial) coding procedures [21] to extract relevant data.

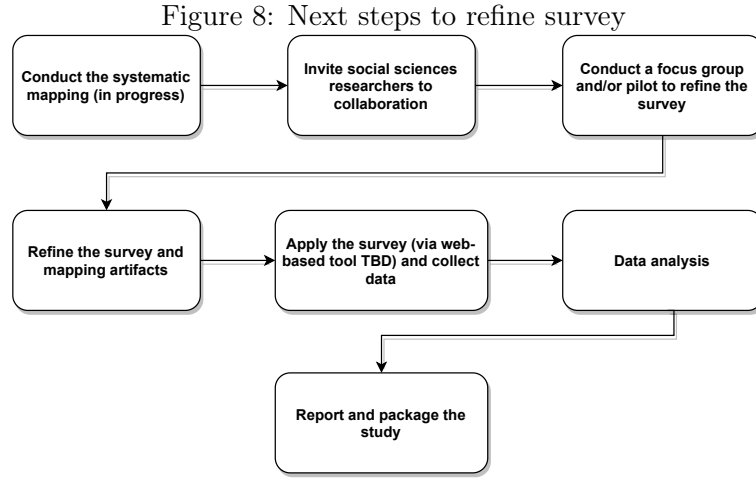
### 4.4 Current Status and Next Steps

In this section we presented our preliminary survey protocol. Currently, we are conducting the data extraction from the candidate set (see Table 1) in the systematic mapping.

Table 8: Survey questionnaire

<i>ID</i>	<i>Questions</i>
Demographic questions	
Q1	What is your current academic degree in social sciences (B.Sc., M.Sc., Ph.D.)?
Q2	How long (in years) do you hold a social sciences bachelor's degree (such as in psychology)?
Q3	How long (in years) have you been researching in the field of psychometrics?
Q4	How long (in years) have you employed psychometric instruments in practice?
RQ2: How the participants perceive the adoption of psychometric instruments regarding personality in SE research? Given the set of psychometric instruments $[x_1, x_2, \dots, x_n]$ used in purpose in SE research $[y_1, y_2, \dots, y_n]$ , please answer:	
Q5	Using the instrument $X$ with the purpose $Y$ in SE research is consistent with its original proposal in social sciences. (strongly agree, agree, neutral, disagree, strongly disagree). Please detail your answer:
Q6	Using the instrument $X$ can support the SE researcher to achieve the purpose $Y$ . (strongly agree, agree, neutral, disagree, strongly disagree) Please detail your answer:
Q7	Instrument $X$ is easy to apply to achieve the purpose $Y$ (strongly agree, agree, neutral, disagree, strongly disagree) Please detail your answer:
Q8	Analyzing and interpreting the result of using instrument $X$ with the purpose $Y$ is easy for a non-social science researchers. (strongly agree, agree, neutral, disagree, strongly disagree) Please detail your answer:
Q9	The application process of the psychometric instrument $X$ with the purpose $Y$ is consistent (strongly agree, agree, neutral, disagree, strongly disagree). Please detail your answer:
Q10	Are there any recommendations to improve the negative points?

The next steps for this survey can be seen in Figure 8.



- **Invite social sciences researchers to collaboration:** we intend to discuss with social sciences researchers about: better definition of the sample; execution of a focus group and/or pilot and data collection agenda.
- **Conduct a focus group and/or pilot study to refine the survey:** with a defined agenda, we plan to conduct a focus group and/or pilot process in a subgroup of respondents. We aim obtain feedback to make refinements on the survey instrument.
- **Refine the survey and mapping artifacts:** with the previous step already executed and discussed, if possible, in collaboration with the social sciences researchers, we will redesign the survey instrument.
- **Apply the survey and collect data:** with a schedule defined, we intend to apply the survey.
- **Data analysis:** the data collected from the survey will be analyzed with the support of the statistical techniques briefly discussed (open to changes).
- **Report and package the study:** finally, we intend to report the results of our study and make the artifacts available in order to make them auditable and replicable.

## 5 Conclusions and Next Steps of The Study

As presented in subsections 3.5 and 4.4, this study is at an initial step. Thus, we cannot draw major discussions and conclusions yet. The next steps direct us to finish the data extraction for the next seven years (2014 - 2020) according to the mapping protocol's inclusion and exclusion criteria (present in Table 1) and thus conduct data extraction as described in the Data Extraction Form (see Table 2). As soon as possible, and with a considerable amount<sup>9</sup> of information extracted, we intend to refine the survey protocol

---

<sup>9</sup>To be discussed among the authors.



with social sciences researchers as discussed in subsection 4.4. Threats to the study's validity need to be strongly discussed among the authors and were not discussed in this ongoing study.

## References

- [1] R. Feldt, R. Torkar, L. Angelis, and M. Samuelsson, “Towards Individualized Software Engineering: Empirical Studies Should Collect Psychometrics,” in *Proceedings of the 2008 International Workshop on Cooperative and Human Aspects of Software Engineering*, CHASE '08, (New York, NY, USA), pp. 49–52, ACM, 2008. event-place: Leipzig, Germany.
- [2] P. Lenberg, R. Feldt, and L. G. Wallgren, “Behavioral software engineering: A definition and systematic literature review,” *Journal of Systems and Software*, vol. 107, pp. 15–37, Sept. 2015.
- [3] D. Graziotin, X. Wang, and Pekka Abrahamsson, “The Affect of Software Developers: Common Misconceptions and Measurements,” May 2015.
- [4] D. Graziotin, F. Fagerholm, X. Wang, and P. Abrahamsson, “What happens when software developers are (un)happy,” *Journal of Systems and Software*, vol. 140, pp. 32–47, June 2018.
- [5] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Berlin Heidelberg: Springer-Verlag, 2012.
- [6] S. Cruz, F. Q. B. da Silva, and L. F. Capretz, “Forty years of research on personality in software engineering: A mapping study,” *Computers in Human Behavior*, vol. 46, pp. 94–113, May 2015.
- [7] S. McDonald and H. M. Edwards, “Who should test whom?,” *Communications of the ACM*, vol. 50, pp. 66–71, Jan. 2007.
- [8] M. Usman and N. M. Minhas, “Use of Personality Tests in Empirical Software Engineering Studies A Review of Ethical Issues,” in *Proceedings of the Evaluation and Assessment on Software Engineering*, EASE '19, (New York, NY, USA), pp. 237–242, Association for Computing Machinery, Apr. 2019.
- [9] D. Graziotin, F. Fagerholm, X. Wang, and P. Abrahamsson, “Consequences of Unhappiness While Developing Software,” in *Proceedings of the 2Nd International Workshop on Emotion Awareness in Software Engineering*, SEmotion '17, (Piscataway, NJ, USA), pp. 42–47, IEEE Press, 2017. event-place: Buenos Aires, Argentina.
- [10] J. Michell, *Measurement in Psychology: A Critical History of a Methodological Concept*. Cambridge University Press, June 1999. Google-Books-ID: oN1cvjpDQeQC.
- [11] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic mapping studies in software engineering,” in *Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering*, EASE'08, (Swindon, GBR), pp. 68–77, BCS Learning & Development Ltd., June 2008.
- [12] K. Petersen, S. Vakkalanka, and L. Kuzniarz, “Guidelines for conducting systematic mapping studies in software engineering: An update,” *Information and Software Technology*, vol. 64, pp. 1–18, Aug. 2015.

- [13] B. Kitchenham, “Guidelines for performing Systematic Literature Reviews in Software Engineering, version 2.3,” Tech. Rep. EBSE-2007-01, Keele University and University of Durham, 2007.
- [14] E. Mourão, M. Kalinowski, L. Murta, E. Mendes, and C. Wohlin, “Investigating the Use of a Hybrid Search Strategy for Systematic Reviews,” in *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pp. 193–198, Nov. 2017.
- [15] E. Mendes, C. Wohlin, K. Felizardo, and M. Kalinowski, “When to update systematic literature reviews in software engineering,” *Journal of Systems and Software*, vol. 167, p. 110607, Sept. 2020.
- [16] C. Wohlin, E. Mendes, K. R. Felizardo, and M. Kalinowski, “Guidelines for the search strategy to update systematic literature reviews in software engineering,” *Information and Software Technology*, vol. 127, p. 106366, Nov. 2020.
- [17] P. Garner, S. Hopewell, J. Chandler, H. MacLehose, E. A. Akl, J. Beyene, S. Chang, R. Churchill, K. Dearness, G. Guyatt, C. Lefebvre, B. Liles, R. Marshall, L. M. García, C. Mavergames, M. Nasser, A. Qaseem, M. Sampson, K. Soares-Weiser, Y. Takwoingi, L. Thabane, M. Trivella, P. Tugwell, E. Welsh, E. C. Wilson, and H. J. Schünemann, “When and how to update systematic reviews: consensus and checklist,” *BMJ*, vol. 354, July 2016. Publisher: British Medical Journal Publishing Group Section: Research Methods & Reporting.
- [18] V. Garousi and J. M. Fernandes, “Highly-cited papers in software engineering: The top-100,” *Information and Software Technology*, vol. 71, pp. 108–128, Mar. 2016.
- [19] S. S. J. O. Cruz, F. Q. B. d. Silva, C. V. F. Monteiro, P. Santos, I. Rossilei, and M. T. d. Santos, “Personality in software engineering: Preliminary findings from a systematic literature review,” in *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*, pp. 1–10, Apr. 2011.
- [20] A.-W. Harzing, “Publish or Perish,” 2020.
- [21] K. Stol, P. Ralph, and B. Fitzgerald, “Grounded Theory in Software Engineering Research: A Critical Review and Guidelines,” in *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*, pp. 120–131, May 2016.
- [22] D. I. K. Sjøberg, T. Dybå, B. C. D. Anda, and J. E. Hannay, “Building Theories in Software Engineering,” in *Guide to Advanced Empirical Software Engineering* (F. Shull, J. Singer, and D. I. K. Sjøberg, eds.), pp. 312–336, London: Springer, 2008.
- [23] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, “Requirements Engineering Paper Classification and Evaluation Criteria: A Proposal and a Discussion,” *Requir. Eng.*, vol. 11, pp. 102–107, Dec. 2005.
- [24] J. Linåker, S. M. Sulaman, R. Maiani de Mello, and M. Höst, “Guidelines for Conducting Surveys in Software Engineering,” 2015. Publisher: [Publisher information missing].

- [25] S. Wagner, D. Mendez, M. Felderer, D. Graziotin, and M. Kalinowski, “Challenges in Survey Research,” in *Contemporary Empirical Methods in Software Engineering* (M. Felderer and G. H. Travassos, eds.), pp. 93–125, Cham: Springer International Publishing, 2020.
- [26] F. D. Davis, “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,” *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989. Publisher: Management Information Systems Research Center, University of Minnesota.

## Annex A List with the final set of included studies (ongoing)

- [S91] Varona, D., Capretz, L.F., Piñero, Y., 2011. Personality types of Cuban software developers. *Global Journal of Engineering Education* 13, 5.
- [S92] Omar, M., Syed-Abdullah, S.-L., Mohd Hussin, N., 2011. Developing a Team Performance Prediction Model: A Rough Sets Approach, in: Abd Manaf, A., Zeki, A., Zamani, M., Chuprat, S., El-Qawasmeh, E. (Eds.), *Informatics Engineering and Information Science, Communications in Computer and Information Science*. Springer, Berlin, Heidelberg, pp. 691–705. [https://doi.org/10.1007/978-3-642-25453-6\\_58](https://doi.org/10.1007/978-3-642-25453-6_58)
- [S93] Salleh, N., Mendes, E., Grundy, J., 2011. The effects of openness to experience on pair programming in a higher education context, in: 2011 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE T). Presented at the 2011 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE T), pp. 149–158. <https://doi.org/10.1109/CSEET.2011.5876082>
- [S94] Alexander, P., Pieterse, V., Lotriet, H., 2011. A Comparison Of Computing And Non-computing Students' Personalities Based On The Five-Factor Model, in: *ECIS 2011 Proceedings*. Presented at the European Conference on Information Systems (ECIS).
- [S95] Monika Mahindra, K. BharathiKumari, 2011. Personality and interest pattern of men in different occupations at Escorts Limited, Faridabad. *Global Journal of Research in Management* 1.
- [S96] Raza, A., UlMustafa, Z., Capretz, L., 2011. Personality Dimensions and Temperaments of Engineering Professors and Students – A Survey. *Journal of Computing* 3, 13–20.
- [S97] Varona, D., Capretz, L.F., 2011. Comparing Cuban and Brazilian software engineers. *World Transactions on Engineering and Technology Education* 9, 5.
- [S98] Martínez, L.G., Castro, J.R., Licea, G., Rodríguez-Díaz, A., Alvarez, C.F., 2021. Knowing Software Engineer's Personality to Improve Software Development. Presented at the 6th International Conference on Software and Data Technologies, pp. 99–104.
- [S99] Yilmaz, M., OConnor, R.V., 2012. Towards the Understanding and Classification of the Personality Traits of Software Development Practitioners: Situational Context Cards Approach, in: 2012 38th Euromicro Conference on Software Engineering and Advanced Applications. Presented at the 2012 38th Euromicro Conference on Software Engineering and Advanced Applications, pp. 400–405. <https://doi.org/10.1109/SEAA.2012.62>
- [S100] Gómez, M.N., Acuña, S.T., Genero, M., Cruz-Lemus, J.A., 2012. How Does the Extraversion of Software Development Teams Influence Team Satisfaction and Software Quality?: A Controlled Experiment. *International Journal of Human Capital and Information Technology Professionals (IJHCITP)* 3, 11–24. <https://doi.org/10.4018/jhcitp.2012100102>
- [S101] Sfetsos, P., Adamidis, P., Angelis, L., Stamelos, I., Deligiannis, I., 2012. Investigating the Impact of Personality and Temperament Traits on Pair Programming: A Controlled Experiment Replication, in: 2012 Eighth International Conference on the Quality of Information and Communications Technology. Presented at the 2012 Eighth International Conference on the Quality of Information and Communications Technology, pp. 57–65. <https://doi.org/10.1109/QUATIC.2012.36>
- [S102] Amir Asil, Hasan Asil, Mohsen Asil, 2012. The Effect of Programmers on Software Project Management Based Upon Personality Five Traits Theory. *IJETAE* 2, 3.
- [S103] Stylianou, C., Andreou, A.S., 2012. A Multi-objective Genetic Algorithm for Software Development Team Staffing Based on Personality Types, in: Iliadis, L., Maglogiannis,

- I., Papadopoulos, H. (Eds.), *Artificial Intelligence Applications and Innovations, IFIP Advances in Information and Communication Technology*. Springer, Berlin, Heidelberg, pp.
- [S104] Rehman, M., Mahmood, A.K., Salleh, R., Amin, A., 2012. Mapping job requirements of software engineers to Big Five Personality Traits, in: 2012 International Conference on Computer Information Science (ICCIS). Presented at the 2012 International Conference on Computer Information Science (ICCIS), pp. 1115–1122.
- [S105] Stylianou, C., Gerasimou, S., Andreou, A.S., 2012. A Novel Prototype Tool for Intelligent Software Project Scheduling and Staffing Enhanced with Personality Factors, in: 2012 IEEE 24th International Conference on Tools with Artificial Intelligence. Presented at the 2012 IEEE 24th International Conference on Tools with Artificial Intelligence, pp. 277–284. <https://doi.org/10.1109/ICTAI.2012.45>
- [S106] Omar, M., Syed-Abdullah, S.-L., Hussin, N.M., 2012. eTiPs: A Rule-based Team Performance Prediction Model Prototype. *Procedia Technology, First World Conference on Innovation and Computer Sciences (INSODE 2011) 1*, 390–394.
- [S107] Montequín, V.R., Balsera, J.V., Fernández, J.M.M., Nieto, A.G., 2012. Using Myers-Briggs Type Indicator (MBTI) as a Tool for Setting up Student Teams for Information Technology Projects. *Journal of Information Technology and Application in Education 7*.
- [S108] Mair, C., Martincova, M., Shepperd, M., 2012. An Empirical Study of Software Project Managers Using a Case-Based Reasoner, in: 2012 45th Hawaii International Conference on System Sciences. Presented at the 2012 45th Hawaii International Conference on System Sciences, pp. 1030–1039. <https://doi.org/10.1109/HICSS.2012.96>
- [S109] Kartelj, A., Filipović, V., Milutinović, V., 2012. Novel approaches to automated personality classification: Ideas and their potentials, in: 2012 Proceedings of the 35th International Convention MIPRO. Presented at the 2012 Proceedings of the 35th International Convention MIPRO, pp. 1017–1022.
- [S110] Gorla, N., Chiravuri, A., Meso, P., 2013. Effect of personality type on structured tool comprehension performance. *Requirements Eng 18*, 281–292.
- [S111] Koroutchev, K., Acuña, S.T., Gómez, M.N., 2013. The Social Environment as a Determinant for the Impact of the Big Five Personality Factors and the Group's Performance. *International Journal of Human Capital and Information Technology Professionals (IJHCITP) 4*, 1–8. <https://doi.org/10.4018/jhcitp.2013010101>
- [S112] Panagiotis Sfetsos, Panagiotis Adamidis, Lefteris Angelis, Ioannis Stamelos, Ignatios Deligiannis, 2013. Heterogeneous Personalities Perform Better in Pair Programming: The Results of a Replication Study. *SQP 15*, 4–15.
- [S113] Bishop, D., 2013. Personality Theory as a Predictor for Agile Preference, in: MWAIS 2013 Proceedings. Presented at the MWAIS conference.
- [S114] Kanij, T., Merkel, R., Grundy, J., 2013. An empirical study of the effects of personality on software testing, in: 2013 26th International Conference on Software Engineering Education and Training (CSEE T). Presented at the 2013 26th International Conference on Software Engineering Education and Training (CSEE T), pp. 239–248.
- [S115] Radhakrishnan, P., Kanmani, S., 2013. Improvement of programming skills using pair programming by boosting extraversion and openness to experience. *IJTCS 4*, 13.
- [S116] Kasturi Dewi Varathan, Li Thing Thiam, 2013. Mining Facebook in Identifying Software Engineering Students' Personality and Job Matching, in: ACSET2013 Conference Proceedings. Presented at the The Asian Conference on Society, Education and

Technology, Osaka, Japan, pp. 278–279.

[S117] Abdul Rehman Gilal, Mazni Omar, Kamal Imran Sharif, 2013. Discovering Personality Types and Diversity Based on Software Team Roles, in: Proceedings of the 4th International Conference on Computing and Informatics. Presented at the International Conference on Computing and Informatics (ICOICI), Sarawak, Malaysia, Universiti Utara Malaysia, p. 6.

[S118] Karapıçak, Ç.M., Demirörs, O., 2013. A Case Study on the Need to Consider Personality Types for Software Team Formation, in: Woronowicz, T., Rout, T., O’Connor, R.V., Dorling, A. (Eds.), Software Process Improvement and Capability Determination, Communications in Computer and Information Science. Springer, Berlin, Heidelberg, pp. 120–129. [https://doi.org/10.1007/978-3-642-38833-0\\_11](https://doi.org/10.1007/978-3-642-38833-0_11)

[S119] McEwan, T., McConnell, A., 2013. Young people’s perceptions of computing careers, in: 2013 IEEE Frontiers in Education Conference (FIE). Presented at the 2013 IEEE Frontiers in Education Conference (FIE), pp. 1597–1603.

[S120] Varona, D., Capretz, L.F., Raza, A., 2013. A multicultural comparison of software engineers. *WIETE* 11, 5.

[S121] Thiti Phiriyayotha, Siriluck Rotchanakitumnuai, 2013. Data Warehouse Implementation Success Factors and the Impact of Leadership and Personality on the Relationship between Success Factors. *Journal of Business and Economics* 4, 9.

[S122] Luse, A., McElroy, J.C., Townsend, A.M., DeMarie, S., 2013. Personality and cognitive style as predictors of preference for working in virtual teams. *Computers in Human Behavior* 29, 1825–1832. <https://doi.org/10.1016/j.chb.2013.02.007>

[S123] Bazelli, B., Hindle, A., Stroulia, E., 2013. On the Personality Traits of StackOverflow Users, in: 2013 IEEE International Conference on Software Maintenance. Presented at the 2013 IEEE International Conference on Software Maintenance, pp. 460–463.

[S124] Martínez, L.G., Licea, G., Rodríguez, A., Castro, J.R., Castillo, O., 2013. Using MatLab’s fuzzy logic toolbox to create an application for RAMSET in software engineering courses. *Computer Applications in Engineering Education* 21, 596–605.