The Journal of Systems & Software 200 (2023) 111654

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Secondary studies on human aspects in software engineering: A tertiary study[☆]

Elijah Zolduoarrati [*], Sherlock A. Licorish, Nigel Stanger

Department of Information Science, University of Otago, Dunedin, New Zealand

ARTICLE INFO

Article history:
Received 27 April 2022
Received in revised form 16 December 2022
Accepted 20 February 2023
Available online 28 February 2023

Keywords:
Tertiary study
Secondary study
Human aspects
Quality assessment
Software engineering

ABSTRACT

Context: This study compiles the evidence reported on the human aspects of software engineering in view of providing a comprehensive catalogue of human aspects that have been examined.
Objective: To summarise the existing systematic literature on human aspects in software engineering.
Method: This study employs published tertiary research guidelines to investigate secondary studies published between 1940 and 2021.
Results: We identified 67 secondary studies concentrating on 16 different human aspects research categories, including Agile, Economic Factors, Environmental Factors to Productivity, Success, and Teams. Several trends reveal the topics that have received the least (e.g., "software engineer controllers") and most (e.g., "individual human aspects") attention in research. Outcomes show that the number of secondary studies on human aspects in software engineering continues to rise when compared to other software engineering topics, despite experiencing a significant drop in 2020. Many secondary studies implemented established guidelines, especially those published in scholarly journals. While there is variance in the quality of published secondary studies, the average quality score across the investigated studies was 3.09 out of 4.0, with journal-published studies and one thesis having higher quality than conference and workshops papers. Specific institutions are also more central to the publication of secondary studies. Furthermore, there has been noteworthy advancement in the consideration of human aspects across the domain. Finally, we discovered several relationships among human aspects investigated. For example, the "Industry" subject of investigation is strongly correlated with the "Theoretical" study type.
Conclusion: The overview provided by this study allows researchers and practitioners to gain familiarity with the current state of research on human aspects, helping researchers to identify gaps for further study and allowing practitioners to discover high-quality, evidence-based approaches to harness the power of human aspects in software engineering.

1. Introduction

Human aspects such as expertise, knowledge, and aptitude play a significant role in software engineering and the software development process, and can directly influence project outcomes (Guveyi et al., 2020). The work regimen has changed over the years due to job requirements being more specific and the high expectations that the employer articulates. Human aspects have started gaining research and industrial interest due to their importance and direct influence on various software engineering dimensions, such as investigating productivity periods where employees excel in their work (Meyer, 2018); work environment conditions (Fagerholm et al., 2014); gender and culture diversity within teams (Catolino et al., 2019; Lee et al., 2011); and other factors that may influence employees' performance. In fact, human aspects are held to influence all areas of software development, from requirements engineering to deployment and maintenance. For instance, improving communication, knowledge sharing, and documentation may enhance members' expertise and positively influence the implementation process, leading to better quality outcomes and more efficient maintenance (Guveyi et al., 2020). Human aspects in software engineering are the influential factors or characteristics that come into play when practitioners form teams to create high quality software. These include education and cognition, soft skills (e.g., communication, leadership and teamwork), practitioners' interest (e.g., preferences for specific processes), extrinsic and intrinsic values such as rewards and motivation (or demotivation) (Guveyi et al., 2020;

☆ Editor: Kelly Blincoe.

* Corresponding author.
E-mail addresses: elijah.zolduoarrati@otago.ac.nz (E. Zolduoarrati), sherlock.licorish@otago.ac.nz (S.A. Licorish), nigel.stanger@otago.ac.nz (N. Stanger).

https://doi.org/10.1016/j.jss.2023.111654

Table 1
Systematic tertiary study research questions.

RQ#	Research question	Motivation
RQ1	What commitment has been invested to gather evidence in software engineering human aspects and the discussed domains?	This question aims to identify the main domains that attracted the highest attention from researchers as well as to highlight knowledge gaps needing more attention.
RQ2	What is the profile of the most influential and active secondary study researchers in the software engineering human aspects domain?	This question aims to investigate the most influential groups of researchers in the discipline, whether other researchers and organisations are also investigating the area, and how their research has been disseminated worldwide. This will help orient future scholars to the work done by such researchers, saving effort and time when there is a need to understand specific human aspects.
RQ3	What are the quality levels of the secondary studies conducted in the software engineering human aspects domain?	This question will aid in identifying quality variations in the research conducted within the discipline and provide recommendations regarding the factors contributing to high quality research.
RQ4	How are various dimensions of human aspects studies related, and what predicts study quality?	This question aims to investigate and identify significant trends within the discipline and whether it is receiving enough attention from researchers, as well as patterns of relations between the human aspect dimensions and study quality.

Fernández-Sanz and Misra, 2011). Thus, understanding human aspects and their direct association to various software development processes could lead to teams achieving stable, reliable, high-quality products (Fernández-Sanz and Misra, 2011).

The primary objective of this study is to summarise what has been done within the sphere of human aspects in software engineering, to catalogue the body of evidence and identify existing knowledge gaps. In so doing, we followed relevant guidelines on the conduct of a tertiary study (Kitchenham et al., 2010). This work provides multiple contributions to software engineering research and practice. From a theoretical research perspective, this work catalogues secondary studies on human aspects in software development, offering an overall lens that bring together the scattered evidence on the subject. Practitioners may explore our findings for guidance in support of their team management and strategies for enhancing human aspects during software development. In fact, our study is particularly relevant given the increasing emphasis on diversified software engineering team environments, where talent is frequently outsourced. Therefore, understanding the influential aspects that may affect team performance is crucial as this knowledge will enable managers and other individuals involved in the team formation process to understand team dynamics, enabling them to tap into their members' hidden talents or avoid undesirable traits, which may lead to better returns. This study also provides methodological guidance, recommendations, and guidelines for those who want to carry out similar tertiary research to advance software engineering theory and practice. We provide an extended taxonomy, as well as a spreadsheet containing human aspects classification in software engineering in our replication package (Zolduoarrati et al., 2022), offering an avenue for human aspects researchers to navigate and understand the most frequently studied domains and topics as well as those needing more attention. Furthermore, this study provides grounded evidence about the most influential and active researchers on human aspects of software engineering over the years. In conclusion, this paper hopes to summarise existing systematic literature regarding various human aspects within the software engineering domain (refer to Table 1 for the research questions).

The remainder of this paper is organised as follows. Section 2 provides the background for our research. Section 3 presents the methods, including the research questions, the search and selection process, inclusion and exclusion criteria, quality criteria, and data extraction. Section 4 provides the results, while the limitations of the research are presented in Section 5. Implications of the study are discussed in Section 6, and we provide concluding remarks in Section 6.

2. Background

Several software engineering researchers have concentrated on diverse software development processes and methods over the years (Zolduoarrati et al., 2022; Wirth, 2008; Sjoberg et al., 2007; Shaw, 2009; Scacchi, 2002). This diversity is, in part, due to the need for software development practitioners to respond to consumer demands. Developing software is complex, going beyond processes and technology. This complexity stems from human interaction between developers and consumers, between developers, and between organisations. Furthermore, many software failures can be linked to human aspects (Unterkalmsteiner et al., 2011). Thus, human aspects are an important part of software development, as they heavily influence the development process.

The software engineering field comprises interdependent individuals acting collectively to achieve positive production quality and group productivity outcomes. Software development is an essentially collaborative team activity (Hazzan and Tomayko, 2004) with a spectrum of complications and uncertainty about which tools and methods to employ (Walz et al., 1993; Sfetsos et al., 2006; Boehm, 2006; Licorish et al., 2009a; Siddiqui and Hussain, 2006), the ideal timeframe for their employment, and who should use them to maximise success rates (Sfetsos et al., 2006; The Standish Group, 1994; Boehm and Turner, 2003a,b; Chin, 2004).

Explanations for project successes and failures remain elusive despite continuous endeavours to improve software development practices (Licorish et al., 2009b). This has driven the focus on promoting a better comprehension of human aspects activities within software development and their role in advancing software development performance (Sfetsos et al., 2006; Acuña et al., 2009; Beranek et al., 2005; Capiluppi et al., 2007; Capretz and Ahmed, 2010; Cunha et al., 2009; Rajendran, 2005). Given the large failure rate of software development projects (The Standish Group, 1994, 2001), investigating and summarising human aspects topics such as interaction and collaboration has become increasingly essential (Coram and Bohner, 2005; Chang and Ehrlich, 2007) to find solutions to push the wheel of development forward and revolutionise current software development frameworks and approaches (Sfetsos et al., 2006; Acuña et al., 2009; Rajendran, 2005; Cataldo et al., 2006; Damian et al., 2007, 2010; Ocker and Fjermestad, 2008). Hence, there is an increasing body of research and evidence dedicated to human aspects and their impact on software development outcomes (Cataldo and Ehrlich, 2012; De Vries et al., 2006; Ehrlich and Cataldo, 2012; Hayes, 2003; Howison et al., 2006; Licorish and MacDonell, 2012; Herbsleb et al., 2001; Abreu and Premraj, 2009; Al-Rawas and Easterbrook, 1996). In particular, empirical evidence has indicated that diversity (social, behavioural, and cultural) aspects are among the most crucial human aspects influencing software development (Walz et al., 1993).

Certain software development uncertainties are explicitly and consistently associated with human aspects, such as group coordination and cohesion, while others appear to be related more to technical issues (Sfetsos et al., 2006; Cataldo and Ehrlich, 2012; Bannerman, 2008; Ropponen and Lyytinen, 2000; Schmidt et al., 2001; Zwikael and Ahn, 2011; Denning, 2017; Sach et al., 2011; Sharma and Kaulgud, 2011; Zhou and Mockus, 2011). Upon closer inspection, however, even the technical issues appear to inevitably channel human aspects, such as inadequate documentation produced by practitioners reporting a defect or a bug. A growing research corpus notes that human aspects such as communication and coordination influence various software development practices. Understanding such factors will lead to recommendations for enhancing quality, project administration, and overall software development processes (Cataldo et al., 2006; Damian et al., 2007, 2010; Ocker and Fjermestad, 2008). The literature has also given significant recognition to software practitioners' behavioural personality concerns, trust, and emotions, providing recommendations for effective management of these characteristics during collaborative activity (Feldt et al., 2010; Gallivan, 2001; Karn and Cowling, 2006; Wynekoop and Walz, 2000; Zheng et al., 2002). A systematic synthesis of available literature on human aspects will allow for a better understanding of the state of the scholarship in this domain.

Kitchenham (2004) label systematic reviews as a scheme that aids scholars to recognise, evaluate, and describe available research linked to a particular subject, discipline, question, or phase. Systematic reviews and mapping studies, hereinafter shall be referred to as *secondary studies* or *secondary reviews (SRs)*. These are studies aimed at investigating primary published research papers associated with research questions or review topics, while tertiary reviews are scholarly research carried out on published secondary research (Glasziou et al., 2000). Tertiary reviews have recently gained traction among researchers in the software engineering community (Petersen et al., 2008; Dybå et al., 2006; Hannay et al., 2007; Kampenes et al., 2007), as they allow scholars to synthesise information on a particular topic in a systematic way.

Petersen et al. (2008) describes systematic literature reviews as a pre-determined investigation to inspect research questions, clarify their methodologies and context, and thoroughly examine their outcomes. Reasons for conducting a secondary study include collecting contemporary evidence concerning a discipline or technique, compiling empirical information such as the advantage of employing a particular methodological aspect and its restrictions, and introducing further avenues of investigation by identifying research (Kitchenham, 2004).

A secondary study of good quality uses a pre-determined search scheme to evaluate the usefulness and composition of the obtained articles, which is necessary to overcome bias and deliver good scientific contributions (Zheng et al., 2002; Kampenes et al., 2007). An additional advantage of engaging in a secondary study is reaching a broad conclusion depending on numerous requirements and frameworks (Kitchenham, 2004; Kitchenham and Charters, 2007a).

Carrying out a secondary study is a demanding task requiring much effort and work (Kitchenham, 2004; Petersen et al., 2008; Kitchenham and Charters, 2007a). Scholars must carefully outline the exclusion and inclusion criteria tailored to the questions of the study. It further employs reviewing mechanisms and a distinctive extraction procedure to determine what information must be collected from the chosen primary or secondary research to evaluate their quality (Kitchenham, 2004).

There is a solid research corpus discussing human aspects in software engineering. Our review of the literature shows several primary and secondary studies on the subject outlining its maturity in past decades (Unterkalmsteiner et al., 2011; Petersen et al., 2008; Jorgensen and Shepperd, 2006; Hall et al., 2009; Dybå et al., 2007). Despite this, we are not aware of any study aimed at holistically gathering and cataloguing the evidence across all human aspects (Mujtaba et al., 2008; Svahnberg et al., 2010; Dyba and Dingsoyr, 2009; Kitchenham et al., 2007). For instance, Hall et al. (2009) examined work on motivation in software engineering published between 1980 and 2006, while Dybå et al. (2007) analysed and reviewed existing collective pair programming literature. Unterkalmsteiner et al. (2011) assessed and evaluated factors affecting the software development process by reviewing 148 articles published between 1991 and 2008, and Jorgensen and Shepperd (2006) conducted a systematic cost estimation review by examining 304 published papers in 76 journals. Although these secondary studies examined and discussed specific human aspects over a particular period, when we searched relevant literature across 12 computer science-related databases (refer to Section 3.2), we were unable to identify a tertiary systematic review that groups under one umbrella all the secondary human aspects studies since the inception of the software engineering discipline in 1940. We found only one tertiary study by Hoda et al. (2017) that analyses and discusses the Agile methodology, which is just one of the sixteen human aspects considered in the current work (see Section 4). A comprehensive tertiary study offers a roadmap for future human aspects researchers to see the most frequently studied domains and which topics require more attention, thereby saving time and resources. The study will update existing secondary research while also providing pointers and recommendations for future researchers interested in conducting tertiary work in software development practice (Sjoberg et al., 2007; Shaw, 2009).

This tertiary study aims to summarise the majority of human aspects in software engineering discipline by reviewing published secondary studies. As an illustration, one of the sixteen human aspects categories captured in this study is 'Individual Human Aspects' (refer to Section 4 for more details), where several intrinsic factors may affect an individual in the software development process. As also captured by Zheng et al. (2002), one such intrinsic factor that motivates a software engineer is trust. Hence, if a software engineer stopped trusting their teammates, they may opt into being a 'solo player' and disassociate themselves from teamwork, which can affect the quality of the produced software.

The sample utilised for this study consisted of systematic literature reviews and mapping studies (referred to as SRs) to ensure good quality and rigour (Dybå and Dingsøyr, 2008). Whilst Kitchenham and Charters' Kitchenham and Charters (2007a) guidelines for systematic literature reviews were published in 2004, this tertiary study examined 148 SRs on software engineering human aspects published between 1940 and 2021 to confirm that no informal SRs were conducted prior to the birth of the aforementioned guidelines. That said, our findings revealed that 2004 was indeed the year when such studies started to gain attention. More digital databases were included than other prior studies, as will be described in the next section. Moreover, this review's broader inclusion and exclusion criteria allowed for reviewing conference proceedings, workshops, and experience reports. We focused on a broader software engineering context of human aspects, since most of the prior SRs focus on a particular factor of human aspects in a narrow software development niche, such as human judgment in cost estimation (Jorgensen and Shepperd, 2006), the investigation of software practitioners' motivations (Hall et al., 2009; França et al., 2011; Beecham et al., 2008), and evaluating and measuring software process improvements (Unterkalmsteiner et al., 2011), among others (Dybå et al., 2007; Varona et al., 2012; Jia et al., 2016).

3. Methods

Our tertiary review employed Kitchenham and Charters' measures and guidelines (Kitchenham et al., 2010; Kitchenham, 2004; Hanssen et al., 2011; Cruzes and Dybå, 2011) and the research questions were developed based on the tertiary model and guidelines (Kitchenham and Charters, 2007b) outlined by Kitchenham et al. (2010). While we provide all relevant details in the manuscript, some elements are too large to fit in their entirety within a printed page. We therefore provide full and exhaustive details in our replication package (Zolduoarrati et al., 2022). Regarding software and tools, in the data collection and preprocessing stages, Microsoft Excel was used along with EndNote to manage the initial set of bibliographies and references. Afterwards, SPSS and Stata were used to conduct data analysis. For graphical rendering, we employed Microsoft Excel and diagrams.net (formerly draw.io). These processes are discussed in the following subsections.

3.1. Systematic tertiary study research questions and motivations

The research questions of this systematic tertiary study, including the motivation behind their implementation, are described in Table 1. RQ1 seeks to identify the most discussed domains in scholarship on human aspects, which will help researchers identify gaps that need to be addressed. RQ2 seeks to identify the prominent researchers and organisations in the field, as well as the most popular research venues, which will enable scholars and practitioners to quickly identify some of the best and most relevant research on human aspects in software development. RQ3 looks at the quality of the identified tertiary studies, revealing the factors that lead to high-quality research.

Quality levels determine the degree of excellence within existing systematic literature in software engineering, providing an overview for future studies to fill in the highlighted flaws. By employing a quality scale, the benefits are not only derived for this research alone, but also for future studies to contribute towards and bolster low-quality works. Finally, RQ4 aims to identify trends within the discipline and to determine what predicts study quality. Table 2 shows the summary of different measures and techniques employed to answer the research questions (RQ1 to RQ4).

3.2. The search method

The search process used online databases that are known to index the software engineering and computer science-related literature (Hoda et al., 2017): IEEE Xplore, ACM, Springerlink, Science Direct, ISI Web of Science, Scopus, Wiley, Citeseerx, Proquest, EBSCOhost, Emerald, and Scirus (refer to Search in the replication package (Zolduoarrati et al., 2022) for a list of database links). Previous software engineering SRs have shown that these provide excellent coverage of available software engineering literature (Kitchenham and Charters, 2007b; Hoda et al., 2017). We searched for all related SR publications from January 1940 to August 2021 within these online databases. The search string composition we adopted was necessarily adjusted to match the search conditions of various digital databases (e.g., IEEE Xplore, Web of Science, and Springer demanded slightly adjusted strings due to the query terms they employ). The search was applied to the title, abstract, and keywords fields to acquire a sufficient sample to answer our research questions.

We initially also intended to include Google Scholar search results, which produced about 16,000 records. However, Google Scholar does not provide a proper filtration mechanism to improve the search as other digital databases do. Also, examining the first five pages revealed that many of the returned results were irrelevant to software engineering and our research questions (refer to Search in the replication package (Zolduoarrati et al., 2022) for Google Scholar screenshots). The search results were instead dominated by human–computer interaction (HCI), medical, and psychological disciplines. Therefore, the Google Scholar database was dropped, as was done when the same challenge presented itself in previous studies (Meldrum et al., 2017; Mohanani et al., 2018).

Due to the breadth of our work, we decided to devise primary and secondary terms to search for SRs studying all three aspects: human aspects, software engineering, and secondary study (refer to Search in the replication package (Zolduoarrati et al., 2022) for detail on search strings), following Hoda et al.'s 2017 approach. We initially formatted the search string as (''human factors'') AND (''systematic review'' OR ''secondary study'' OR ''mapping study''). However, the results exceeded 40,000, including studies from many unrelated disciplines. Therefore, we limited our investigation to the software engineering field only. We also noted that researchers have used different morphemes and terms to investigate similar constructs. For example, Chagas et al. (2015) used the term ''human factors'' to investigate the impact of humans on agile software projects, while Hidellaarachchi et al. (2021) used the term ''human aspects'' to investigate the effects of humans on the requirements engineering process. Both studies review primary studies to investigate the relationship between humans and a specific software engineering context. Thus, the search expression had to be modified accordingly to accommodate those cases. That is, the search string for ''human'' was conducted using secondary terms of ''human aspects'', ''human factors'', ''human issues'' or ''human errors''. Furthermore, the search string for ''software'' was conducted using secondary terms ''software engineering'' or ''software development''. Finally, the search string for ''systematic'' was conducted using the secondary terms ''systematic review'', ''systematic literature review'', ''systematic map'' or ''mapping study''. Each database has its own search and filtration mechanisms. Therefore, one database's search formatting may not apply to another (e.g., ''Metadata'' in IEEE point to the same entity as ''Abstract'' in ACM). Despite this disparity, we have used similar combinations as seen in Table 3.

Reflecting on the work by Zhang et al. (2011), we carried out an automated search where we evaluated the search results based on our initial set of papers as benchmark. If the result from the automated search was not satisfactory (e.g., discusses about game development instead of software engineering in general), the search string was refined. Otherwise, the search result would be considered for inclusion. According to Cruz et al. (2015), the quality and efficiency of search strategies employed in secondary studies present significant issues in any evidence-based discipline. Such concepts are usually operationalised by two indexes: sensitivity and precision. The authors define sensitivity as the ratio between the number of related studies retrieved and the total number of existing relevant studies in the literature. In contrast, precision is the ratio between the number of related studies retrieved and the total number of studies retrieved. An optimal strategy should achieve the highest possible values for both indexes. To do so, we used inclusion and exclusion criteria followed by backwards snowballing techniques to achieve the highest possible sensitivity and precision. By doing backwards snowballing, we anticipated that studies may be related and may likely be cited. Given that our search period starts from 1940 and was current at August 2021, after conducting our extensive search, doing backwards snowballing should uncover most of the relevant papers. In fact, when we conducted backwards snowballing, we found many studies that were returned by our initial

Table 2
Different measures and techniques for each RQ.

RQ#	Measures	Techniques
RQ1	Quantitative analysis	Mathematical frequency was conducted to examine the number of SRs each year, as well as which human aspect domains gained more attention in general. In return, this provides insights for future studies to fill gaps in the human aspects space.
	Time-based analysis	Time span between SRs were analysed to examine studies' timeline. This was done to examine variation of timeline across all SRs (e.g., several SRs span as long as 10 years, while others cover as short as 2 years), and provide robustness for the results of further analysis.
	Thematic analysis	Thematic analysis identified patterns in the examined SRs, with the goal to report repeated patterns. Top-level categories were first classified, where each category consists of subcategories, and further subdivided into topics and subtopics.
RQ2	Quantitative analysis	For the SRs, quantitative analysis was done to highlight notable patterns in the number of citations, authors, institutions and countries. From such analyses, we were able to derive top-cited SRs, most influential authors, as well as countries and institutions with the most SRs.
RQ3	Descriptive statistics	Frequency and mean quality per publication venue (conference, journal, thesis, workshops) were calculated to determine which type of publication yielded the highest quality, as well as which venues invited the most SRs.
	Pearson's correlation analysis	Pearson's correlation analysis was conducted to investigate correlation between quality score and different categories of study type, subject of investigation (human aspects), and methods. On a 5% significance level, the null hypothesis was that the correlation was zero and the alternate hypothesis was that the correlation was not zero.
RQ4	Pearson's correlation analysis	Similar to RQ3, Pearson's correlation analysis was done to inspect relationships across various dimensions of human aspects studies (e.g., whether the "empirical" study type correlated with "industry" subjects). The null hypothesis was that the correlation between two variables was statistically not different from zero, while the alternate hypothesis was that the correlation between two variables was statistically different from zero.
	OLS multilinear regression	Ordinary least squares (OLS) multilinear regression is widely used to investigate the determinants of a particular variable, due to its simplicity and applicability. The model was constructed using quality score as a dependent variable, while the independent variables were the number of primary studies covered, number of years covered, number of authors, total citations, total institutions, total countries, total number of topics (human aspects), total study types, total subjects of investigation, and total methods.
	Breusch–Pagan heteroskedasticity test	The Breusch–Pagan test investigates whether heteroskedasticity issues (i.e., non-constant variance of residuals) are present in regression. If residuals are heteroskedastic (i.e., OLS assumption is violated), then it will be difficult to trust the standard errors of the OLS estimates. The null hypothesis from the test is that residuals are homoskedastic (constant variance).

Table 3
String formatting for searching.

Primary terms	Secondary terms
Human	*("human aspects" OR "human factors" OR "human issues" OR "human errors")*
	AND
Software	*("software engineering" OR "software development")*
	AND
Systematic	*("systematic review" OR "systematic literature review" OR "systematic map" OR "systematic mapping" OR "mapping study")*

search. Informal skimming of a few studies for relevant forward citations returned studies already in our pool. Therefore, we did not see the need for forward snowballing.

3.3. Study selection

Kitchenham (2004) suggested that a selection protocol is essential for an efficient and effective secondary study, as this phase could be influenced by researcher bias. Reducing such bias can be achieved if several researchers collectively formulate a protocol supported by a pilot study evaluating the protocol's reliability. This approach has been recommended by Brereton et al. (2007) to enhance the efficacy, reliability, and validity of the different research stages while acknowledging time and scheduling constraints. As shown in Table 4, the publication exclusion and inclusion criteria were based on the selection rules, resulting in a consistent database for the data extraction stage.

This protocol and its criteria emerged from thorough discussions between three software engineering researchers without any conflicts, thereby requiring no further protocol modification. The following selection protocol was formally expressed and subsequently applied to the records collected from the search phase, increasing the relevancy of search results. Finally, due to the researchers' language limitations, papers written in languages other than English were disregarded.

Duplicated studies (E.1) were identified by manually checking for similarities in titles, author names, and abstract, and a similar process was carried out to identify papers that were of insufficient breadth (E.2). Secondary studies based on industrial surveys, developer comments, unclear systematic format, and non-empirical studies (E.3) were excluded due to several reasons including lack of access, patent contents, or because they were not conducted empirically using scientific guidelines (Gregory et al., 2015; Stavru, 2014; Abrahamsson et al., 2003). Studies published outside of our university's subscription (which included most of the major databases) were excluded as "inaccessible reviews" (E.4). Such studies were locked behind a paywall that required a significant payment to obtain the full text. Therefore, due to time constraints, the breadth of our work and the high quality of the databases queried, we did not contact authors and consequently decided to exclude those studies.

The paper selection process is shown in Fig. 1 and followed a similar approach to Hoda et al. (2017) and Kitchenham et al. (2010). The first step in the process was the initial search that produced unfiltered results using the search strings, resulting in a total of 1,563 publications. Applying the inclusion criterion reduced this to 148, and applying the exclusion criteria filtered out another 92 to give 56.

Table 4
Summary of inclusion/exclusion selection mechanism.

Inclusion	**(I.1)** The SRs must be related to human aspects in software engineering/development topics. **(I.2)** The studies must be written in the English language.
Exclusion	**(E.1)** Duplicated studies. **(E.2)** SRs that appear to be only an abstract, work-in-progress paper, poster, or short paper (four or fewer pages not covering the full breadth of a secondary study). **(E.3)** Secondary studies based on industrial surveys, developer comments, unclear systematic format, or non-empirical. **(E.4)** Inaccessible studies.

Fig. 1. Paper selection process flow.

Ten random papers (approximately 5%) were then piloted by two researchers (the first author and a trained rater). Individuals involved in the piloting process were presented with guidelines for applying, marking, and comprehending the inclusion/exclusion criteria and examples of the type of publication that needs to be included or excluded to improve the reliability and consistency of this phase. Both raters were in total agreement, leading the primary researcher to rate the rest of the publications, resulting in 56 studies.

At this stage, we decided to perform a secondary search mechanism, applying the backward snowballing technique (Wohlin, 2014) to identify additional related studies. This was conducted by iteratively and exhaustively reviewing titles from the list of references within the chosen studies through four iterations (Mohanani et al., 2018). In each iteration, studies of interest were assessed against the exclusion and inclusion guidelines, leading to better precision and improvement in the publication sample, which resulted in 68 records. The next step was to perform a quality assessment step where records of inferior quality (equal to zero on the quality scale, see Section 3.4) were excluded, resulting in a final set of 67 SRs available for meta-analysis (comprising 45 systematic literature reviews and 12 mapping studies across 38 conference papers, 25 journal articles, one thesis, and three workshops). The entire selection process was discussed among the researchers to enhance the reliability of the final sample and minimise potential biases, including but not limited to search string creation, digital database selection, inclusion/exclusion criteria determination, quality assessment, randomly analysing some of the selected papers, and finally thematic analysis to yield the final sets of topics (refer to Section 4.1. for additional details).

3.4. Quality assessment

The quality of the SRs was assessed based on the guidelines described in Kitchenham et al. (2010). These were based on the York University Database Abstracts of Reviews of Effects (DARE) criteria (Centre for Reviews and Dissemination, 2016), leading to the formulation of four evaluation questions to assess the quality of each SR as presented in Table 5. The selected SRs were evaluated using a point system to classify how well publications satisfied the quality criteria questions: Yes (Y) = 1 point, Partial (P) = 0.5 points, No (N) = 0 points. This point-based assessment was first introduced by Kitchenham (2004) and Hoda et al. (2017). Considering the breadth of our work, we perceived that the quality criteria used in these studies were sufficient for our quality evaluation. Furthermore, the creation of new criteria (as well as their evaluation) would take a significant amount of time, and so, we felt that this would be outside the scope of the work.

The total quality score was determined by adding up the four distinct criteria scores. Thus, the complete quality score for individual studies ranged between 0 (i.e., satisfying none of the criteria) and 4 (i.e., fully satisfying all four criteria). These are similar conditions to those employed to assess research quality in tertiary systematic review studies by Kitchenham et al. (2010, 2009). To verify the quality scheme's reliability and validity, ten random papers were selected for piloting. Similar to the previous inclusion/exclusion phase, two researchers (the first author and a trained rater) performed the quality assessment, and discrepancies between raters were discussed and resolved through consensus. Cohen's Kappa coefficient (Cohen, 1960), a measure of inter-rater reliability, indicated an agreement level between raters of 0.79 for the first iteration, and 0.853 for the second iteration (refer to Quality in the replication package (Zolduoarrati et al., 2022) for details), which is considered a strong level of agreement (Cohen, 1960; McHugh, 2012). The first author went on to analyse the quality of the remaining papers. The full list of quality agreement, as well as an example of Cohen's Kappa calculation may also be found within Quality in the replication package (Zolduoarrati et al., 2022).

The SR quality scores are presented in Fig. 2. The figure shows that almost 70% of the SRs had a quality score of 3.0 or higher. Approximately 50% of the SRs had a quality score of 3.5 or higher. This points to the mostly excellent quality of the SRs identified in this study.

3.5. Data extraction and analysis process

The data extraction was done using a structured extraction form in Microsoft Excel to capture all required information for additional synthesis. The extracted data included bibliographic information (title, publication year, SR guidelines,[1] publication venue), SR study parameters (SR topics, number of primary studies, primary publication periods, number of citations, online databases searched, and quality score), and primary public information (study types, investigation subjects, and discussed methods).

[1] Guidelines provide information about the type of structure or format that the secondary study follows. For example, some of the secondary studies are formatted are based on Kitchenham and Charters' structure (Kitchenham and Charters, 2007a) while others are based on Petersen et al. (2015) or the mix of two types, etc.

Table 5
Quality assessment questions and measures.

Q#	Quality question	Quality assessment measures
Q1	Are the SR's inclusion and exclusion conditions reported appropriately?	• The inclusion/exclusion measures are not adequately defined and cannot easily be implied ⇒ N (no) • The inclusion measures are implicit ⇒ P (Partial) • The inclusion/exclusion measures are explicitly and adequately described in the paper ⇒ Y (yes)
Q2	Was the research reasonably explored to cover all related studies?	• The researcher(s) have examined a maximum of two online libraries or utilised restricted access journal collections ⇒ N (no) • The researcher(s) have examined three to four online libraries with no additional search procedures, or defined a search employing restricted access journals or conference proceedings collections ⇒ P (Partial) • The researcher(s) examined four or more online libraries, incorporated additional search procedures, and acknowledged all publication venues discussing the subject of interest through the appropriate citations ⇒ Y (yes)
Q3	Have reviewers evaluated the quality and validity of the incorporated studies?	• The quality of the data was not considered in the research, or the quality evaluation of individual papers was not attempted ⇒ N (no) • The quality of the data was obtained but not defined appropriately in the research, or the quality evaluation of individual papers was partially included ⇒ P (Partial) • The researcher(s) quality criteria have been outlined explicitly and extracted from all primary studies ⇒ Y (yes)
Q4	Were the primary data or studies represented appropriately?	• The researcher(s) have not cited information related to the extracted primary studies ⇒ N (no). • Only a summary of the primary studies is presented, such as grouping papers into classes, but it is quite complex to associate individual studies with every class ⇒ P (Partial) • Data summaries can be traced back to related papers due to sufficient information about each paper being present ⇒ Y (yes)

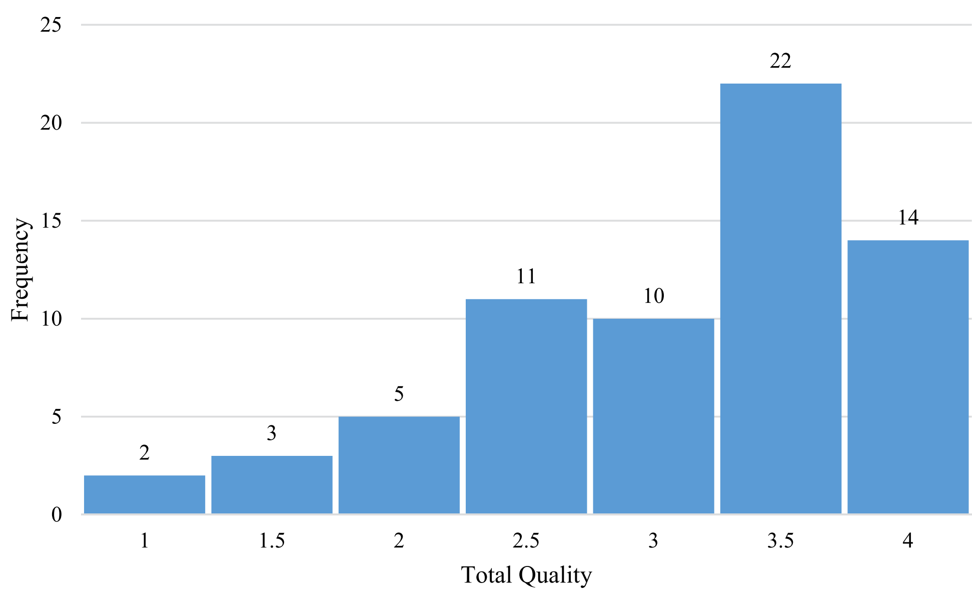

Fig. 2. Quality measurement scores of SRs.

The extracted records were grouped into domains to answer the RQs (Cohen, 1960) through a categorisation process, then a statistical analysis of related numerical data was carried out (Braun and Clarke, 2006). For example, answering RQ1 involved descriptive analysis of the number of SRs and the type of domains covered over the years, to observe whether the discipline is receiving enough attention from researchers and what are the most and least popular human aspects domains discussed over the years. To answer RQ2, we examined the most active or influential individuals, organisations, and publication venues in human aspects SRs to help future scholars more deeply probe the body of work, thereby saving effort and time. To answer RQ3, we assessed the quality levels of the included SRs to check whether there have been efforts in the software engineering human aspects domain towards achieving good quality using specific SR guidelines. To answer RQ4, we performed Pearson's correlation analysis to explore various trends in human aspects topics, study types, subjects of investigation, methods used, and guidelines cited. This will enable future scholars to either use similar methods or to innovate new methodological approaches to find related literature gaps, and to check whether there is improvement using specific SR guidelines. Table 6 summarises the data extracted from all included SRs.

4. Results and discussion of research questions

The 67 SRs examined in this study are summarised in Table 7. For each SR, we identified the total quality score, year of publication, type of publication venue, number of primary studies covered by the review, range of years covered, SR guidelines cited, and discussed domains. These reviews represent primary studies covering an overall period of 59 years. Among the 67 SRs, 38 studies were published in conference proceedings, 25 in journals, one as a thesis, and the remaining three as workshop papers. Most of the SRs (about 88%) followed Kitchenham's guidelines (Kitchenham, 2004; Kitchenham and Charters, 2007a; Centre for Reviews and Dissemination, 2016), while others cited Beecham et al. (2008), Petersen et al. (2015), Dyba et al. (2007), Webster and Watson (2002), and Brereton et al. (2007), among others. Two SRs Yaseen et al. (2015), Gomes and Marczak (2012) did not explicitly cite guidelines but still fulfilled the quality criteria described in Section 3.4.

Table 6
Secondary studies extracted information entries.

Primary and secondary extracted information	
Bibliographic information	• Title • Publication year • SR guidelines • Publication Venue (Conference/Journal/Workshop/Thesis)
SR study parameters	• SR topics • Number of primary studies • Primary publication periods • Number of citations • Online databases searched • Quality score
Primary publication information	• Study types (empirical, theoretical, or experimental, or other study types) • Subjects of investigation (industry, non-industry, both, or other subject of investigation) • Discussed methods (experiment, case study, reports, ethnography, survey, literature review, solution proposal, dataset, interview, observation, multiple methods approach, or other methods)

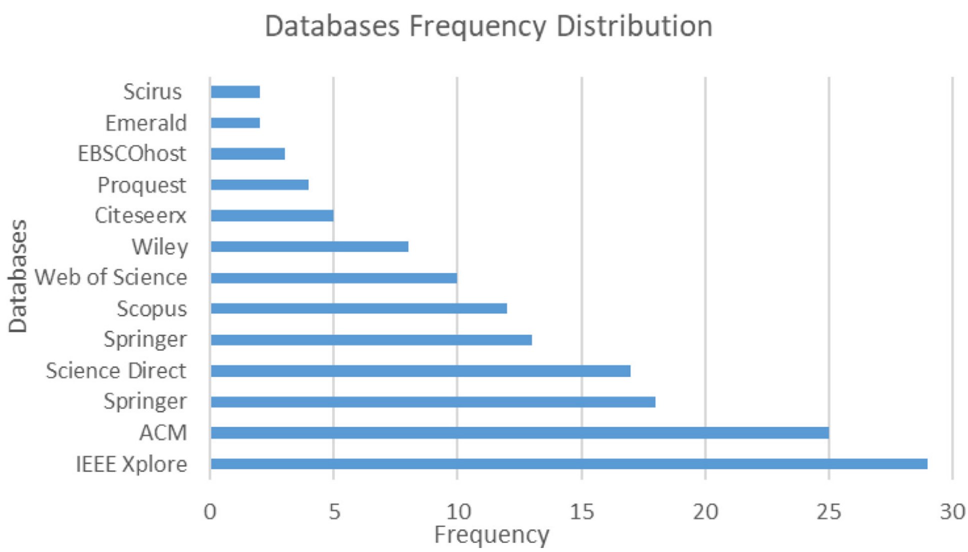

Fig. 3. Frequency distribution of publication databases used in SRs.

Analysis of the data extracted found that the Individual Human Aspects domain was the most investigated, appearing in 17.4% of SRs. This was followed by Team (11.9%); Software Development (Workflow) and Management (both 11%); Success Factors (9.2%); Agile, Social and Innovative, and Knowledge Sharing (all 7.3%); Other (6.4%); Experience and Productivity (4.6%); Economic Factors and Technical Human Aspects (2.8%); Environmental Factors (1.8%); and Soft Skills and Software Engineer Controllers (both 0.9%).

Fig. 3 shows the number of SRs that conducted their searches using each online publication database. IEEE Xplore and ACM emerged as the most preferred online databases to conduct human aspects related searches, while Emerald and Scirus were the least used.

Table 7 reveals that the SRs comprised 38 conference papers (56.7%), 25 journal articles (37.3%), one thesis (1.5%), and three workshops (4.5%). In terms of study types, Table 8 shows that empirical studies were the most frequent type of primary study found by the SRs. The total number of primary studies under empirical type was 1,129, followed by other study types (303), theoretical study (140) and experimental study (14). Furthermore, in terms of subject of investigation, industry was the most frequent focus of investigations across all venue types. Table 8 reveals that 505 primary studies focussed on industry, followed by other subject of investigation (503 primary studies), non-industry (385 primary studies), and both (204 primary studies).

Finally, SRs across all venue types revealed that primary studies often used multiple methods approach (see Table 8). There were 918 primary studies that used multiple methods approach, followed by those that used only case studies (447 primary studies), questionnaires (420 primary studies), and other methods (382). On the other hand, Table 8 shows that observation was used the least (11 primary studies).

4.1. Research question 1

The first research question is: *What commitment has been invested to gather evidence in software engineering human aspects and the discussed domains?*

We found that 67 SRs were published between 2007 and 2021 on various human aspects topics. Twelve of these were mapping studies, while the remaining were systematic literature reviews. Fig. 4 shows the number of SRs that were published each year.

The trend indicates a growing interest in investigating human aspects in the Software Engineering field since 2007. The number of SRs published peaked in 2019, suggesting that interest associated with human aspects in this field may be decreasing or that the main topics of interest have been covered. However, the COVID-19 pandemic may also have played a role in reduced publication outcomes, since interactions between scholars were restricted, thus potentially limiting this type of research.

In terms of the number of years covered in each SR, as noted in Table 7, we argue that the sample utilised in the analysis

Table 7
Human aspects in software development SRs published between 2007 and 2021.

ID	Quality	Year	Venue	Studies	Years	Guidelines	Domains
Walia and Carver (2009)	4	2009	Journal	149	1961–2007 (46)	KI	Social and Innovative, Technical Human Aspects, Management, Software Development (Workflow)
Shameem et al. (2018)	3.5	2018	Conference	18	2001–2017 (16)	KI	Individual Human Aspects
Silveira and Prikladnicki (2019)	1	2019	Workshops	221	2002–2018 (16)	PE	Knowledge Sharing, Individual Human Aspects
Matturro et al. (2019)	2.5	2018	Journal	44	1999–2017 (18)	KI	Soft Skills
Vishnubhotla et al. (2018)	3.5	2018	Conference	16	2006–2015 (9)	KI	Individual Human Aspects
Lenberg et al. (2015)	2.5	2015	Journal	250	1997–2013 (16)	KI	Individual Human Aspects
Garousi et al. (2016)	2.5	2016	Journal	33	1980–2014 (34)	KE, KI, PE	Software Development (Workflow)
da Silva et al. (2010)	4	2010	Conference	54	1998–2009 (11)	BE, KI	Management
Mohanani et al. (2018)	4	2018	Journal	65	1990–2016 (26)	PE, KI	Management, Software Development (Workflow), Other
Anu et al. (2018)	2.5	2018	Journal	38	1990–2006 (16)	KI	Technical Human Aspects
Dybå and Dingsøyr (2008)	4	2008	Journal	36	1998–2005 (7)	CO	Agile
Alzoubi et al. (2016)	3.5	2016	Journal	21	2006–2014 (8)	KI, DY, RO	Social and Innovative
Brito et al. (2020)	3.5	2020	Journal	56	2013–2017 (4)	KI, CH	Social and Innovative
Onoue et al. (2018)	3.5	2015	Journal	90	1970–2010 (40)	KI, CH	Individual Human Aspects
Pedreira et al. (2015)	4	2015	Journal	29	2010–2014 (4)	PE, KI	Software Development (Workflow)
Onoue et al. (2018)	3.5	2018	Journal	78	2013–2017 (4)	KI, CH, PE	Individual Human Aspects
Guveyi et al. (2020)	2.5	2020	Conference	80	2009–2019 (10)	KI	Individual Human Aspects
Pirzadeh (2010)	4	2010	Thesis	67	2000–2010 (10)	KI	Individual Human Aspects, Economic Factors, Management, Software Development (Workflow)
El Bajta and Idri (2020)	2.5	2020	Conference	12	2015–2018 (3)	KI, CH	Economic Factors
Barroso et al. (2017)	3	2017	Conference	21	2003–2016 (13)	KI	Individual Human Aspects
Sharp et al. (2009)	3.5	2008	Journal	92	1980–2006 (26)	KI	Individual Human Aspects
Beecham et al. (2008)	4	2007	Journal	92	1980–2005 (25)	KI	Individual Human Aspects
Ghanbari et al. (2018)	2	2018	Journal	19	1968–2005 (37)	KI, CH	Management
Cruz et al. (2011)	2.5	2011	Conference	42	1970–2010 (40)	KI, CH	Individual Human Aspects
Ciupe et al. (2017)	2	2017	Conference	176	2000–2017 (17)	KI, PE	Agile
Groeneveld et al. (2019)	1.5	2019	Journal	26	1994–2018 (24)	KI	Experience
Schneider et al. (2013)	2.5	2012	Journal	330	1998–2012 (14)	KI	Team, Knowledge Sharing, Individual Human Aspects, Other
Chagas et al. (2014)	3	2014	Conference	34	2001–2013 (12)	KI	Agile
Sánchez-Gordón and Colomo-Palacios (2019)	3.5	2019	Journal	66	2005–2018 (13)	KI, CH	Individual Human Aspects
Hidellaarachchi et al. (2021)	3.5	2021	Journal	74	1997–2019 (22)	KI, CH	Team, Individual Human Aspects, Technical Human Aspects
Hummel et al. (2013)	2.5	2013	Journal	333	1996–2012 (16)	WE, WA	Team, Management, Other
Lenberg et al. (2014)	2	2014	Conference	52	1997–2013 (16)	KI	Individual Human Aspects
Khan et al. (2017)	3.5	2017	Conference	19	1998–2017 (19)	KI, CH	Success Factors
Dutra et al. (2015)	2	2015	Conference	15	2008–2015 (7)	KI	Team, Individual Human Aspects, Management

(continued on next page)

Table 7 (*continued*).

Costa and França (2019)	1	2019	Conference	17	2002–2008 (6)	KI	Environmental Factors
Jabbari et al. (2016)	3	2016	Workshops	44	1969–2015 (46)	KI, BR	Team, Experience, Software Development (Workflow)
Heaton and Carver (2015)	3	2015	Journal	43	1999–2015 (16)	KI, CH	Software Development (Workflow)
Veras et al. (2020)	3	2020	Conference	26	2008–2020 (12)	KI, WR	Team
de Barros Sampaio et al. (2010)	1.5	2010	Conference	60	1975–2010 (35)	KI	Productivity
Sirazitdinov et al. (2020)	2	2020	Conference	125	1970–2020 (50)	BR, KI	Team, Productivity, Success Factors, Knowledge Sharing, Agile, Individual Human Aspects, Experience, Other
Soomro and Salleh (2014)	3.5	2014	Conference	21	1993–2011 (18)	KI, CH	Team
Flores and de Alencar (2020)	3.5	2020	Conference	37	2009–2019 (10)	KI, CH	Software Engineer Controllers
Melo et al. (2012)	2.5	2012	Conference	12	2001–2011 (10)	BE, KI	Agile
Canedo and Santos (2019)	2.5	2019	Conference	57	2009–2019 (10)	KI	Productivity
Komal et al. (2019)	3.5	2019	Conference	29	2010–2018 (8)	KI	Social and Innovative
Fatima et al. (2019)	3.5	2019	Conference	33	2013–2019 (6)	KI, CH	Knowledge Sharing
Ramírez-Mora and Oktaba (2017)	4	2017	Conference	25	2007–2016 (9)	KI	Productivity
Anchundia and Fonseca (2020)	3.5	2020	Journal	40	2002–2018 (16)	KI	Knowledge Sharing, Social and Innovative, Experience
Nazir et al. (2019)	3	2019	Conference	30	2013–2019 (6)	KI, CH	Team, Knowledge Sharing, Other
Senapathi and Srinivasan (2013)	4	2013	Conference	18	2005–2012 (7)	KI, DY	Agile
Sinha et al. (2020)	3.5	2020	Conference	20	2006–2018 (12)	TO	Success Factors
Lima et al. (2019)	3	2019	Conference	53	2001–2018 (17)	KI	Knowledge Sharing, Productivity, Social and Innovative, Management, Software Development (Workflow)
Fitriani et al. (2016)	3.5	2016	Conference	20	2012–2016 (4)	KI, CH	Agile
Shameem et al. (2017)	3.5	2017	Conference	85	1995–1998 (3)	SH, RA	Team
Yaseen et al. (2015)	3.5	2015	Conference	71	2000–2014 (14)	–	Success Factors
Ali and Khan (2014)	4	2014	Conference	111	1994–2014 (20)	KI, CH	Success Factors
Suali et al. (2017)[a]	3	2017	Conference	20	2000–2015 (15)	KI	Success Factors
Anh et al. (2012)[a]	3.5	2012	Conference	28	2005–2011 (6)	KI	Team
Kroll et al. (2013)[a]	3.5	2013	Conference	36	1990–2012 (22)	KI, CH	Team, Success Factors, Social and Innovative
Jia et al. (2016)[a]	4	2016	Workshops	40	2002–2014 (12)	KI	Environmental Factors,
Gomes and Marczak (2012)[a]	1.5	2012	Conference	202	1997–2013 (16)	–	Social and Innovative, Management, Software Development (Workflow)
Abrar et al. (2019)[a]	3	2019	Journal	58	1992–2016 (24)	KI, BO, DY	Success Factors
Uludag et al. (2018)[a]	3	2018	Conference	73	2005–2017 (12)	CR, DY	Agile
Ali et al. (2017)[a]	4	2017	Journal	152	2001–2016 (15)	KI, ST	Success Factors
Durán et al. (2019)[a]	3.5	2019	Conference	26	2016–2017 (1)	KI	Team, Individual Human Aspects, Experience, Management, Software Development (Workflow), Other
Anwar et al. (2019)[a]	4	2019	Journal	42	2010–2017 (7)	KI, CH	Knowledge Sharing
Britto et al. (2014)[a]	4	2014	Conference	24	2001–2013 (12)	KI, CH	Economic Factors

KI = Kitchenham, CH = Charters, PE = Petersen, KE = Keele, BE = Beecham, CO = Cochrane, DY = Dyba, RO = Rowe, WE = Webster, WA = Watson, WR = Wright, BR = Brereton, TO = Tollgate, SH = Shin, RA = Ramesh, CR = Cruzes

[a]Studies included from backwards snowballing.

is comprehensive. The time span covered varies across SRs, as shown in Fig. 5. Some SRs that cover a long time span are Walia and Carver (2009), Garousi et al. (2016), Onoue et al. (2018), Ghanbari et al. (2018), Cruz et al. (2011), Jabbari et al. (2016), de Barros Sampaio et al. (2010), and Sirazitdinov et al. (2020). The shortest time span covered is in Durán et al. (2019). The oldest study covered by the SRs is 1961 in Walia and Carver (2009), while the most recent year included is 2021 in Sanchez-Gordon

Table 8
SR study types, subjects of investigation, and methods.

ID	Study type				Subject of investigation				Method											
	Empirical	Theoretical	Experimental	Other study types	Industry	Non-Industry	Both	Other subject of investigation	Experiment	Case study	Reports	Ethnography	Questionnaire/survey	Literature review	Solution proposal	Dataset	Interview	Observation	Multiple methods approach	Other methods
Walia and Carver (2009)	–	–	–	5	–	–	–	12	–	–	–	–	–	–	–	–	–	–	–	7
Shameem et al. (2018)	–	–	–	3	–	–	–	6	–	–	–	–	–	–	–	–	–	–	–	20
Silveira and Prikladnicki (2019)	–	–	–	19	–	–	–	3	–	–	–	–	–	–	–	–	–	–	–	20
Matturro et al. (2019)	44	–	–	–	–	–	–	14	1	–	–	–	13	12	–	–	6	–	19	–
Vishnubhotla et al. (2018)	16	–	–	–	–	46	–	–	–	–	1	–	12	–	–	–	6	–	5	–
Lenberg et al. (2015)	190	60	13	–	141	46	63	–	–	–	–	–	–	–	–	–	–	–	–	16
Garousi et al. (2016)	27	–	–	–	1	17	15	–	–	1	–	–	–	–	–	–	1	–	1	–
da Silva et al. (2010)	22	20	–	–	–	–	–	18	–	–	11	–	–	1	–	–	–	–	42	–
Mohanani et al. (2018)	47	–	–	–	–	–	–	5	–	–	–	–	–	–	–	–	–	–	–	5
Anu et al. (2018)	–	–	–	14	–	–	–	13	–	–	–	–	–	–	–	–	–	–	–	12
Dybå and Dingsøyr (2008)	36	–	–	–	24	12	–	–	3	24	–	–	4	–	–	–	–	–	5	–
Alzoubi et al. (2016)	21	–	–	–	–	–	–	8	–	–	–	–	–	–	–	–	–	–	–	14
Brito et al. (2020)	31	–	–	–	–	56	–	–	–	–	–	–	–	–	–	–	–	–	–	5
Onoue et al. (2018)	–	–	–	3	27	45	13	–	24	17	–	2	34	–	–	–	–	–	–	20
Pedreira et al. (2015)	29	–	–	–	–	29	–	–	–	–	–	–	–	–	11	–	–	–	18	–
Onoue et al. (2018)	54	22	1	–	–	–	–	11	–	9	–	–	4	–	–	1	–	–	–	13
Guveyi et al. (2020)	–	–	–	19	–	–	–	5	13	–	–	–	34	–	–	28	12	9	–	8
Pirzadeh (2010)	55	17	–	–	67	–	–	–	9	43	–	–	26	–	36	–	–	–	–	3
El Bajta and Idri (2020)	12	–	–	–	–	–	–	7	–	–	–	–	–	–	–	–	–	–	–	20
Barroso et al. (2017)	16	4	–	–	–	–	–	12	–	–	–	–	–	–	–	–	–	–	–	4
Sharp et al. (2009)	–	–	–	6	–	–	–	11	–	–	–	–	–	–	–	–	–	–	–	16
Beecham et al. (2008)	79	10	–	–	–	–	–	8	–	–	–	–	72	–	–	–	–	–	20	–
Ghanbari et al. (2018)	16	–	–	–	4	15	–	–	–	3	–	–	1	–	–	–	–	–	–	–
Cruz et al. (2011)	36	6	–	–	15	20	1	–	14	11	–	1	10	–	–	–	–	–	–	–
Ciupe et al. (2017)	–	–	–	6	–	–	–	13	–	–	–	–	–	–	–	–	–	–	–	16
Groeneveld et al. (2019)	–	–	–	12	–	–	–	15	–	12	–	–	–	6	–	–	–	–	8	–
Schneider et al. (2013)	–	–	–	20	–	–	–	5	–	–	–	–	–	–	–	–	–	–	–	9
Chagas et al. (2014)	–	–	–	8	–	–	–	17	–	–	–	–	–	–	–	–	–	–	–	2
Sánchez-Gordón and Colomo-Palacios (2019)	66	–	–	–	49	12	5	–	42	18	–	1	5	–	–	–	–	–	–	–
Hidellaarachchi et al. (2021)	5	–	–	–	73	23	4	–	5	6	–	–	3	–	–	–	3	2	55	–
Hummel et al. (2013)	–	–	–	2	–	–	–	15	21	128	5	14	26	20	–	–	–	–	119	–

(*continued on next page*)

Table 8 (*continued*).

Lenberg et al. (2014)	–	–	–	13	22	8	17	–	1	28	1	–	29	1	15	–	–	–	–	–
Khan et al. (2017)	5	–	–	–	–	19	–	–	–	–	–	–	–	–	–	–	–	–	–	1
Dutra et al. (2015)	–	–	–	1	9	4	2	–	–	–	–	–	–	–	–	–	–	–	–	18
Costa and França (2019)	–	–	–	12	–	–	–	20	–	–	–	–	–	–	–	–	–	–	–	2
Jabbari et al. (2016)	44	–	–	–	–	–	–	20	–	–	–	–	–	–	–	–	–	–	–	9
Heaton and Carver (2015)	–	–	–	13	–	–	–	3	–	–	–	–	–	–	–	–	–	–	–	5
Veras et al. (2020)	–	–	–	13	–	–	–	1	–	–	–	–	12	–	–	–	4	–	10	–
de Barros Sampaio et al. (2010)	–	–	–	20	–	–	–	16	–	–	–	–	–	–	–	–	–	–	–	10
Sirazitdinov et al. (2020)	–	–	–	7	–	–	–	19	–	4	–	–	15	–	–	–	–	–	–	–
Soomro and Salleh (2014)	21	–	–	–	14	7	–	–	–	–	–	–	–	–	–	–	–	–	–	16
Flores and de Alencar (2020)	–	–	–	13	–	–	–	4	–	–	–	–	–	–	–	–	–	–	–	8
Melo et al. (2012)	–	–	–	13	–	–	–	17	–	–	–	–	–	–	–	–	–	–	–	18
Canedo and Santos (2019)	57	–	–	–	–	–	–	9	–	–	–	–	–	–	–	–	–	–	–	13
Komal et al. (2019)	–	–	–	18	–	–	–	13	–	–	–	–	–	–	–	–	–	–	–	9
Fatima et al. (2019)	4	–	–	–	–	–	–	17	–	4	–	–	–	–	–	–	–	–	–	13
Ramírez-Mora and Oktaba (2017)	5	1	–	–	–	–	–	2	3	12	–	–	6	–	–	–	–	–	4	–
Anchundia and Fonseca (2020)	40	–	–	–	–	–	–	4	–	–	–	–	–	–	–	–	–	–	–	9
Nazir et al. (2019)	–	–	–	17	–	–	–	16	–	–	–	–	–	–	–	–	–	–	–	15
Senapathi and Srinivasan (2013)	18	–	–	–	18	–	–	–	–	2	–	–	6	–	–	–	8	–	–	–
Sinha et al. (2020)	20	–	–	–	–	–	–	20	–	–	–	–	–	–	–	–	–	–	–	6
Lima et al. (2019)	53	–	–	–	–	–	–	14	–	–	–	–	–	–	–	–	–	–	–	20
Fitriani et al. (2016)	20	–	–	–	15	–	5	–	–	–	–	–	–	20	–	–	–	–	–	–
Shameem et al. (2017)	40	–	–	–	–	–	–	8	5	25	–	–	40	9	–	–	6	–	–	–
Yaseen et al. (2015)	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	71	–
Ali and Khan (2014)	–	–	–	–	–	–	–	9	–	–	–	–	–	–	–	–	–	–	111	–
Suali et al. (2017)	–	–	–	–	–	–	–	14	–	–	–	–	–	–	–	–	–	–	20	–
Anh et al. (2012)	–	–	–	–	–	–	–	20	2	6	–	–	13	–	9	–	3	–	–	–
Kroll et al. (2013)	–	–	–	–	–	–	–	19	–	–	–	–	–	–	–	–	–	–	36	–
Jia et al. (2016)	–	–	–	–	–	–	–	4	–	–	–	–	–	–	–	–	–	–	40	–
Gomes and Marczak (2012)	–	–	–	–	–	–	–	14	–	–	–	–	–	–	–	–	–	–	202	–
Abrar et al. (2019)	–	–	–	3	–	–	–	1	–	17	14	–	10	5	–	–	7	–	5	–
Uludag et al. (2018)	–	–	–	–	–	–	73	–	–	–	–	–	–	–	–	–	–	–	73	–
Ali et al. (2017)	–	–	–	20	–	–	–	20	–	62	–	–	37	24	–	–	21	–	–	–
Durán et al. (2019)	–	–	–	4	–	–	–	–	–	–	–	–	–	–	–	–	6	–	20	–
Anwar et al. (2019)	–	–	–	19	26	2	6	–	1	15	–	–	8	8	–	–	–	–	10	–
Britto et al. (2014)	–	–	–	–	–	24	–	–	–	–	–	–	–	–	–	–	–	–	24	–

and Sánchez-Gordón and Colomo-Palacios (2019). Overall, Fig. 5 demonstrates that the sample of SRs included in the analysis is heterogeneous. That is, the sample does not only include SRs that cover short or long periods only, which should make the analysis more accurate. Furthermore, in terms of the number of primary studies covered, the sample of SRs taken in the analysis is also heterogeneous. Two SRs (Schneider et al., 2013, and Hummel et al., 2013) provide the highest number of primary studies at 330 and 333, respectively. The reason these two SRs include such a high number of primary studies may relate to the broad nature of their RQs. For instance, Hummel et al. (2013) is a secondary study that concentrates on behavioural and social aspects in the work activities of software engineers. Schneider et al. (2013) investigated the common conferences and journals in the software engineering field to present an analysis of solutions for the challenges faced by global software engineers. Fig. 5. Years covered for the SRs. shows that the earliest research revolving around human aspects within software development dates to 1961 and was introduced in S1. However, research on human aspects did not gain significant attention until around 1970, as seen in Brito et al. (2020), Ghanbari et al. (2018), Cruz et al. (2011), Jabbari et al. (2016), de Barros Sampaio et al. (2010), Sirazitdinov et al. (2020). Another interesting thing to note is that S65 conducted a secondary study with the least span of years, dating only from 2016 to 2017.

To identify the topic areas covered by the SRs, the authors carried out a thematic analysis, adapting the process from Braun and Clarke (2006), and using Microsoft Excel to support the

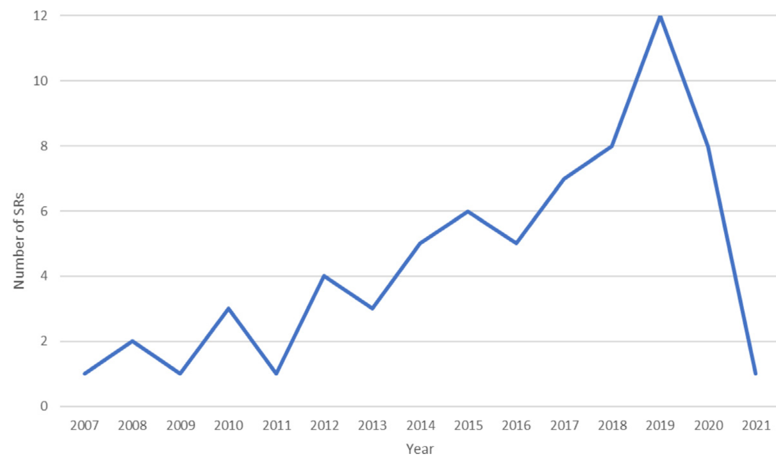

Fig. 4. Number of SRs published each year (2007–2021).

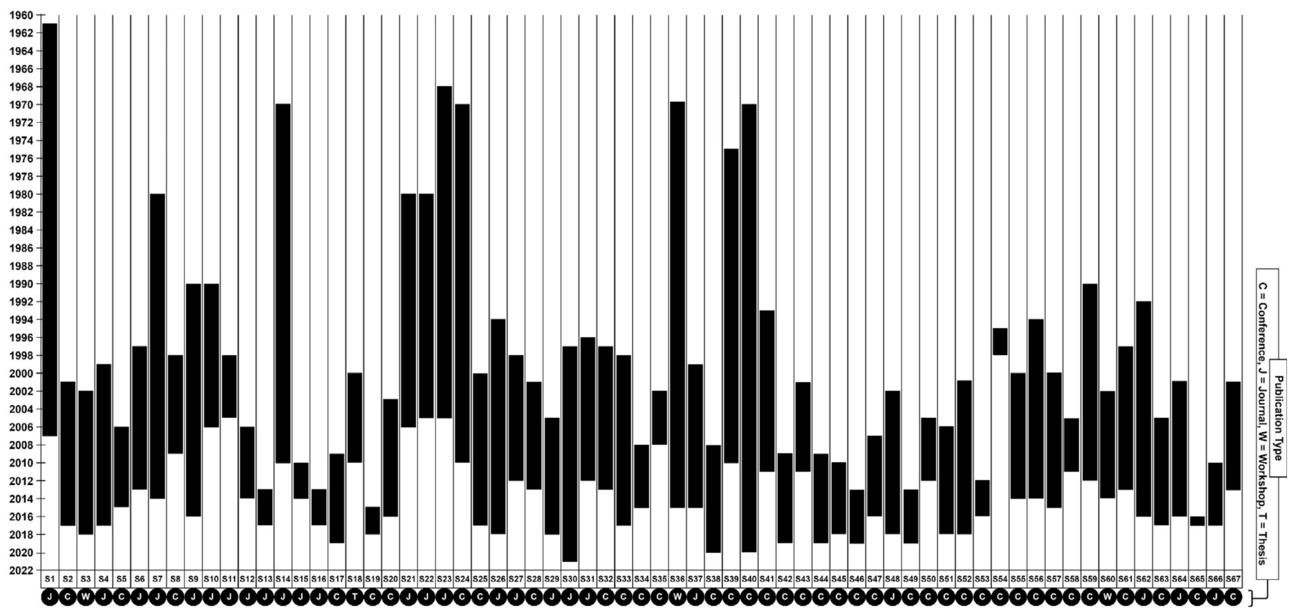

Fig. 5. Years covered for the SRs.

analysis. We initially started by looking for patterns of meaning and issues of potential interest in the data. Next, analysis involved a constant back and forth across the entire set of SRs where we reviewed distinct topics of human aspects in software engineering (taken directly from the studies). Finally, we tally these distinctions to yield our final set of topics. Multiple meetings were conducted to discuss the topic categorisation process as well as the hierarchical structure of the extracted information. Subsequently, the topics catalogue was devised (refer to Human Aspects Taxonomy in the replication package (Zolduoarrati et al., 2022) for a detailed topics catalogue diagram) where we captured sixteen top-level categories from our analysis of the SRs. Furthermore, each category consists of subcategories, and further subdivided into topics and subtopics. Due to the breadth of our data, we have decided to answer RQ1 by deriving patterns from the topics catalogue due to it being too large to display. The authors discussed the categorisation procedure and shared suggestions in a meeting where no significant disagreements in the categorisation procedure were found (as this detail was extracted directly from the papers). The categories were identified based on the specified topics addressed in SRs, making it easier to extract them without significant abstraction and inferencing. For example, Hidellaarachchi et al. (2021) investigated the effect of 'Emotions' as a subcategory within the 'Individual Human Aspects,' category while Sánchez-Gordón and Colomo-Palacios (2019) carried out similar research. They classified emotions into three topics, 'Positive', 'Negative' and 'Neutral', with various subtopics beneath these (e.g., the 'Positive' topic contained 'joy', 'love', 'happiness', and 'excitement', while the 'Negative' topic contained 'anger', 'sadness', and 'fear' subtopics). Thus, we merged the two 'Emotions' subcategories under the 'Individual Human Aspects' category since the former was considered primary. It is important to note that a category usually consists of several subcategories that branch into topics and subtopics. For example, as illustrated in Fig. 6 the "Individual Human Aspects" category consisted of several subcategories including the "Motivators in Software Engineering" and this subcategory branched into several topics including the "Trust/respect" topic that had its own subtopic "External signs of motivated software engineers". The full topic catalogue can be found in our replication package (Zolduoarrati et al., 2022).

Answering RQ1: The sixteen top-level categories identified were Agile, Economic Factors, Environmental Factors, Experience,

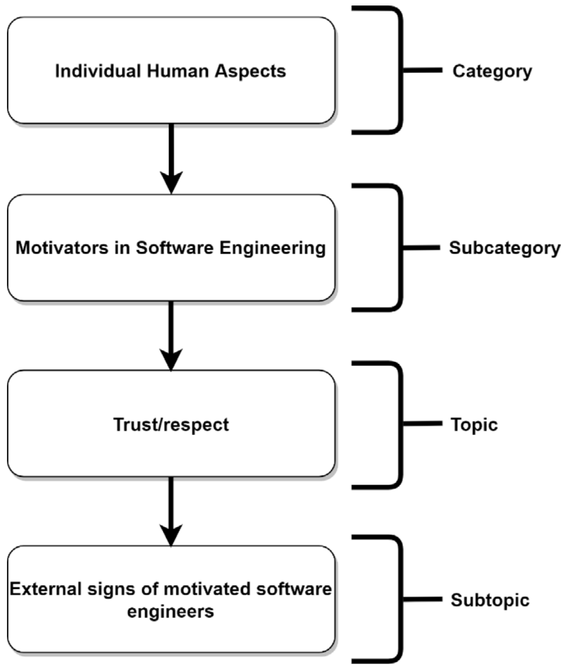

Fig. 6. Human aspects hierarchical structure.

Individual Human Aspects, Knowledge Sharing, Management, Productivity, Social and Innovation, Soft Skills, Software Development (Workflow), Software Engineer Controllers, Success, Team, Technical Human Aspects, and Other. Fig. 7 shows the total number of primary studies underlying the SRs for each category. We found the Success Factors domain had the largest number of studies (2,004) followed by Individual Human Aspects with 1,416, Management with 959, Team with 772, Social and Innovative with 695, and Agile with 663. The least studies, however, are in the Economic Factors domain with 196 studies, and Other with 120, and Technical Human Aspects with 57.

The sixteen top-level research categories are presented in the first column of Table 9. The second column presents a brief description of what was covered by the category. The third column lists the total number of unique SRs under each category (e.g., Individual Human Aspects had the highest number of SRs at 19). Finally, the last column lists the citations supporting each category.

For example, Barroso et al. (2017) and Onoue et al. (2018) investigated personality in software engineering, covering 149 primary studies. A brief description of the SR topics covered under each of the sixteen top-level categories is provided below:

Agile covered 663 primary studies and included several subcategories. One subcategory was 'Large-Scale Agile Development', which was covered in SRs carried out by Chagas et al. (2014), Sirazitdinov et al. (2020), Senapathi and Srinivasan (2013), and Uludag et al. (2018). These SRs identified 11 topics, including various subtopics, and covered 259 primary studies. Another subcategory was 'challenges', which was identified by Fitriani et al. (2016), who performed a secondary study that identified thirty topics related to various challenges facing software engineers during agile software development, from 103 primary studies. Other subcategories that were investigated by Ciupe et al. (2017), Melo et al. (2012), and Dybå and Dingsøyr (2008) included 'Education' (45 studies), and 'Software Engineering - general' (36 studies). 'Information Science' covered 32 studies, 'Agile Manifesto implications' 25 studies, 'Motivators' 23 studies, 'Agile alternative' 18, 'Industrial Applicability' 15, 'Demotivators' 14, 'Mixed technical topics' 13, 'Software models and methods' 7, and 'Training' and 'Agile Mindset' each covered 6 primary studies. The remaining subcategories were covered in five or fewer primary studies.

Economic Factors covered 196 primary studies that included two main subcategories, the 'Identified (cost) drivers' investigated by Britto et al. (2014) (136 studies), and 'Cost Estimation' investigated by Pirzadeh (2010) (60 studies).

Environmental Factors was investigated by Jia et al. (2016) and Costa and França (2019). It covered 295 primary studies which included 37 subcategories, including 'Work environment' with 26 studies, 'Autonomy' with 19 studies, 'Task identity' with 14, 'Communication' with 13, 'Task significance' with 12, and 'Empowerment' with 11 primary studies; while the rest of the subcategories each covered 10 or fewer studies.

Experience was investigated by several authors, including Durán et al. (2019) and Sirazitdinov et al. (2020). It covered 204 primary studies which included thirty subcategories such as 'Program comprehension' with 30 studies, 'Teamwork/dynamics' with 25, 'Bridge the gap' with 14, 'Self-reflection' and 'Conflict resolution' with 13 each, 'Novice programmers' with 11, 'Mentoring' with 10, 'Individual performance' with 9, 'Experience with technology' with 7, 'Working experience/previous projects' with 6 primary studies, and other subcategories each covering 4 or fewer studies.

Individual Human Aspects covered 1,468 primary studies and included 12 subcategories including the 'Behavioural factors' subcategory, which was extensively investigated by Lenberg et al. (2015), covering 371 primary studies with fifty-six topics. This was followed by the 'Motivation' subcategory, which covered 296 primary studies, and the 'Identity' subcategory which covered 217 primary studies, investigated by Silveira and Prikladnicki (2019). The 'Emotions' subcategory was investigated by Sánchez-Gordón and Colomo-Palacios (2019) and covered 162 primary studies. The 'Software engineer personality characteristics' subcategory was investigated by Barroso et al. (2017), Cruz et al. (2015), and Cruz et al. (2011), covering 149 primary studies. 'Human challenges' was investigated by Shameem et al. (2018), covering 105 primary studies while the 'Human capital' subcategory was investigated by Onoue et al. (2018), covering 90 primary studies. There were other subcategories under 'Individual Human Aspects', covering a small number of studies such as 'Human Values' (17 studies), 'Physical Issues' (4), and 'Adaptation/Negotiation' (3).

Knowledge Sharing was extensively investigated by Fatima et al. (2019) and Mohanani et al. (2018), covering 477 primary studies with 19 subcategories including 'Individual Pressures' which was the largest subcategory with 231 primary studies. This was followed by the subcategory 'Organisational Facilitators' with 31 studies, 'Individual Barriers' with 28 studies, 'Organisational Barriers' with 26, 'Technological Facilitators' with 25, 'Cultural Barriers' and 'Technological Barriers' (both with 24 studies), 'Individual Facilitators' (19), 'Cultural Facilitators' (18), 'Geographical Barriers' (16), 'Individual Historical Factors' (10), 'Individual Emotions' (8), while the remaining eight subcategories each covered 4 or fewer primary studies.

Management was investigated by various authors, including da Silva et al. (2010) and Pirzadeh (2010), covering 959 primary studies. It included seven subcategories where the 'Project Management' subcategory was the largest, covering 843 primary studies, followed by 'Software Engineering Measurement' and 'Activity Planning/Configuration Management' (each with 38 studies), 'Resource Allocation' (27), and 'Review and Evaluation' (8).

Productivity covered 364 primary studies and included 34 subcategories that have been examined by various authors, including Ramírez-Mora and Oktaba (2017) and de Barros Sampaio et al. (2010). The largest subcategories were 'Process' and 'Tools' subcategories, each covering 25 primary studies, followed by 'Team skills' (24), 'Programming language' (21), 'Software complexity' (20), 'Coordination and leadership' (19), 'Software size'

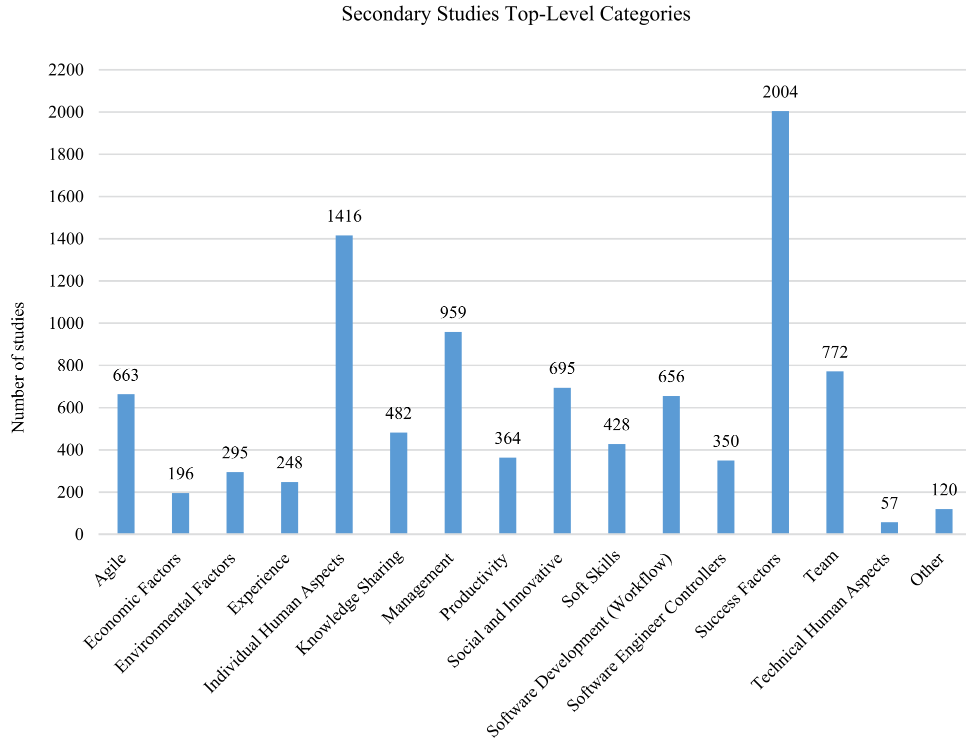

Fig. 7. Total human and social topics per domain.

(19), 'Reuse' (17), 'Software Engineering tasks' (13), 'Requirements stability' (12), 'Communication' (11), 'Motivation' (10), 'Management Quality' (10), 'Knowledge Management tasks' (9), 'Development environment' (9), 'Organisational context' (9), 'Reward System' (8), 'Client participation and experience' (8), 'Balance of member contribution and backup behaviour' (8), 'Amount of Documentation' (7), 'Team organisation' (6), and 'Turnover' (6). The remaining subcategories each covered 5 or fewer primary studies.

Social and Innovative was investigated by various authors, including Vishnubhotla et al. (2018), Kroll et al. (2013), Gomes and Marczak (2012), Walia and Carver (2009), Brito et al. (2020), and Komal et al. (2019). It covered 695 primary studies and included eight subcategories: 'Communication' (376 studies), 'Scope creep' (152), 'Social machines'(135), 'Personal' (8), 'Affective' (5), 'Social' (5), 'Interpersonal' (4), and 'Work ethics' (2 primary studies).

Soft Skills covered 428 primary studies and 30 main subcategories with no other topics, as extensively investigated by Matturro et al. (2019). The included subcategories were 'Communication skills' (40 primary studies), 'Team work' (30), 'Organisational/Planning skills' (24), 'Analytical skills' (24), 'Interpersonal skills' (23), 'Problem solving skills' (21), 'Leadership' (21), 'Autonomy' (19), 'Decision making' (15), 'Initiative' (14), 'Conflict Management' (14), 'Commitment/Responsibility' (13), 'Stress Management' (13), 'Results orientation' (11), 'Time Management' (11), 'Ethics' (11), 'Presentation skills' (10), 'Innovation' (10), 'Creativity' (9), 'Critical thinking' (9), 'Negotiation skills' (9), 'Listening skills' (8), 'Motivation' (8), 'Willingness to learn' (8), while the rest of the subcategories each covered 5 or fewer studies.

Software Development (Workflow) was investigated by various authors such as Pirzadeh (2010), Heaton and Carver (2015), Hidellaarachchi et al. (2021), and Walia and Carver (2009), covering 656 primary studies and 17 subcategories including 'Maintenance' (120 primary studies), 'Requirements' (181), 'Testing' (94), 'Development and Operations' (57), 'Scheduling' (54), 'Software process model and life cycle model' (39), 'Implementation' (31), 'Design' (24), 'Software construction' (16), 'Lifecycle model' (16), 'Verification and Validation' (12) and 'Refactoring' (5). The remaining subcategories each covered 3 or fewer primary studies.

Software Engineer Controllers covered 350 primary studies, and 8 subcategories: 'Competencies' (314 studies), 'Culture' (8), 'Career stage' (8), 'Personality traits' (7), and 'State of IT profession' (5). The rest of the subcategories each covered 4 or fewer studies.

Success Factors was the most investigated category covering 2,004 studies. It was investigated by various authors including Yaseen et al. (2015), Abrar et al. (2019), Ali et al. (2017), Ali and Khan (2014), Sinha et al. (2020), and Khan et al. (2017). This category contains 83 subcategories including 'Communication, Collaboration, and Coordination' (183 studies), 'Organisational proximity' (79), 'Barriers' (78), 'Mutual interdependence and shared values' (76), 'Quality production' (70), 'Mutual trust' (66), 'Flexible Service Level Agreements' (65), 'Effective and timely communication' (64), 'Organisational proximity' (58), 'Lack of effective and proper way of communication' (58), 'Organisational changes' (56), 'Long term commitments' (56), 'Joint management infrastructure' (50), and 'Requirement Management' (50). There were 68 other subcategories that each covered fewer than 50 primary studies.

Team covered 772 studies and was investigated by various authors, including Shameem et al. (2017), Vishnubhotla et al. (2018), Nazir et al. (2019), and Schneider et al. (2013). This particular category contained 52 subcategories including 'Virtual team challenges' (409 studies), 'Professional' (97), 'Dispersion' (52), 'Team Building/Management' (50), 'Communication Issues' (33), 'Team Interaction' (21), 'Team climate' (13), 'Team effectiveness' (12), 'Geographical Distance/Time zone' (8), and 'Team process' (6). The rest of the subcategories each covered 4 or fewer primary studies.

Technical Human Aspects covered 57 studies and was investigated by various authors, including Hidellaarachchi et al. (2021), Anu et al. (2018), and Walia and Carver (2009). This category contained five subcategories including 'Human error' (31 studies)

Table 9
Sixteen top-level categories in human aspects identified in the tertiary study.

Category	Description	Count	Citation(s)
Agile	Agile practices affect software engineers' human and social aspects, such as examining what motivates software developers in agile settings.	8	Melo et al. (2012), Fitriani et al. (2016), Sirazitdinov et al. (2020), Senapathi and Srinivasan (2013), Uludag et al. (2018), Chagas et al. (2014), Dybå and Dingsøyr (2008), Ciupe et al. (2017)
Economic Factors	Cost-related attributes leading to various challenges in the context of software development that influence the way developers behave and conduct their work.	3	Britto et al. (2014), El Bajta and Idri (2020), Pirzadeh (2010)
Environmental Factors	Features such as workspace impacting software practitioners' adaptation to their development environment that affect various social and human aspects such as behaviour, interaction, and decision making.	2	Costa and França (2019), Jia et al. (2016)
Experience	Various factors related to the mental state and proficiency of people engaged in software development, particularly software developers.	5	Sirazitdinov et al. (2020), Jabbari et al. (2016), Anchundia and Fonseca (2020), Durán et al. (2019), Groeneveld et al. (2019)
Individual Human Aspects	Human-centred activities affecting people working closely and effectively with diverse stakeholders, software development team members, and other activities that influence software development such as emotions, motivation, and gender.	19	Silveira and Prikladnicki (2019), Schneider et al. (2013), Hidellaarachchi et al. (2021), Dutra et al. (2015), Sirazitdinov et al. (2020), Durán et al. (2019), Shameem et al. (2018), Vishnubhotla et al. (2018), Lenberg et al. (2015), Onoue et al. (2018), Cruz et al. (2015), Guveyi et al. (2020), Pirzadeh (2010), Barroso et al. (2017), Sharp et al. (2009), Beecham et al. (2008), Cruz et al. (2011), Sánchez-Gordón and Colomo-Palacios (2019), Lenberg et al. (2014)
Knowledge Sharing	Different operations through which knowledge is exchanged among software developers that influence their performance, such as cognitive load, memory and perception.	8	Schneider et al. (2013), Sirazitdinov et al. (2020), Nazir et al. (2019), Silveira and Prikladnicki (2019), Fatima et al. (2019), Anchundia and Fonseca (2020), Lima et al. (2019), Anwar et al. (2019)
Management	Properties related to developers managing group work, developers' communication in groups, multi-cultural group/project management, training staff, developers' personalities and relationship to role assignment, and managing psychological atmosphere in a development environment such as managing developer and staff resistance to using new methods or interacting with different divisions.	10	Walia and Carver (2009), Pirzadeh (2010), Hummel et al. (2013), Dutra et al. (2015), Lima et al. (2019), Gomes and Marczak (2012), Durán et al. (2019), da Silva et al. (2010), Mohanani et al. (2018), Ghanbari et al. (2018)
Other	Various individual factors that have not been categorised or cannot be linked to a particular category (domain) despite the evident link that connects it to the human and social domain of software engineering, such as customer and stakeholder feedback and its effect on the final software product.	6	Mohanani et al. (2018), Schneider et al. (2013), Hummel et al. (2013), Sirazitdinov et al. (2020), Nazir et al. (2019), Durán et al. (2019)
Productivity	Factors influencing practitioners' productivity, performance, and the amount of produced work in their software development environments, such as rewards and mutual support and trust.	5	Sirazitdinov et al. (2020), Lima et al. (2019), de Barros Sampaio et al. (2010), Canedo and Santos (2019), Ramírez-Mora and Oktaba (2017)
Social and Innovative	Properties related to the process of developing and deploying effective solutions to challenges developers face in their development environment, to enhance their creativity and support their social progress.	8	Anchundia and Fonseca (2020), Lima et al. (2019), Kroll et al. (2013), Walia and Carver (2009), Alzoubi et al. (2016), Brito et al. (2020), Komal et al. (2019), Gomes and Marczak (2012)
Soft Skills	Interpersonal skills demonstrating a developer's ability to communicate effectively and build relationships with others in one-on-one interactions and in groups and teams.	1	S4
Software Development (Workflow)	Activities in software development influencing practitioners such as development method chosen by developers and software maintenance.	10	Walia and Carver (2009), Mohanani et al. (2018), Pirzadeh (2010), Jabbari et al. (2016), Lima et al. (2019), Gomes and Marczak (2012), Durán et al. (2019), Garousi et al. (2016), Pedreira et al. (2015), Heaton and Carver (2015)
Software Engineer Controllers	Mixed software engineering drivers that regulate the strength of certain characteristics of developers such as demographic influence, practitioner roles, and personality types.	1	Flores and de Alencar (2020)
Success	Various elements necessary for developers to achieve their mission and goals such as the ability to have a long commitment to a project and to work in a cross-cultural development environment.	9	Sirazitdinov et al. (2020), Kroll et al. (2013), Khan et al. (2017), Sinha et al. (2020), Yaseen et al. (2015), Ali and Khan (2014), Suali et al. (2017), Abrar et al. (2019), Ali et al. (2017)

(*continued on next page*)

Table 9 (*continued*).

Category	Description	Count	Citation(s)
Team	Properties affecting multiple developers working together on a project or maintaining a product, such as self-organising and cross-functional issues.	13	Schneider et al. (2013), Hidellaarachchi et al. (2021), Hummel et al. (2013), Dutra et al. (2015), Jabbari et al. (2016), Veras et al. (2020), Sirazitdinov et al. (2020), Soomro and Salleh (2014), Nazir et al. (2019), Shameem et al. (2017), Anh et al. (2012), Kroll et al. (2013), Durán et al. (2019)
Technical Human Aspects	Technological factors beyond software development that influence software developers, such as misunderstanding how the software interfaces with the rest of the system, or incomplete knowledge about organisational technical policies.	3	Walia and Carver (2009), Hidellaarachchi et al. (2021), Anu et al. (2018)

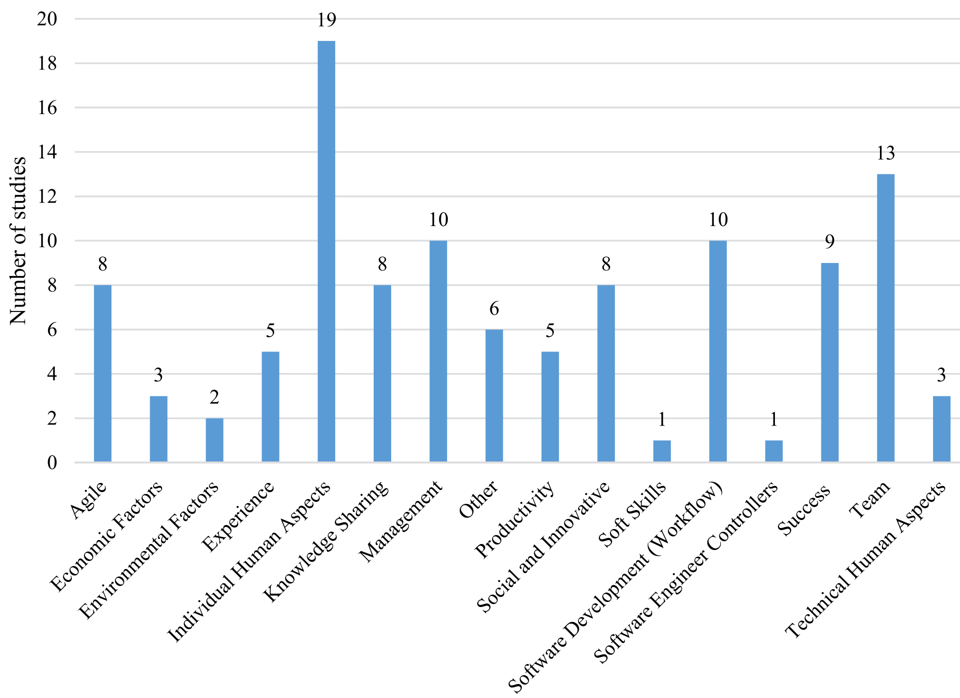

Fig. 8. Total number of SRs per human aspect researched.

and 'Domain Knowledge' (22 studies), while the remaining subcategories were 'Technical capacity of equipment', 'Accessibility', and 'Task workload' each covering 2 or fewer studies.

Other covered 120 primary studies with 21 subcategories, with research carried out by Durán et al. (2019). This category comprised subcategories that did not have a clear association with any of the other fifteen top-level categories. The 'Variety tools' subcategory covered 69 primary studies, followed by 'Reviewer Response' (34 studies), 'Construction' (14), 'Customer' (14), 'Software delivery' (6), 'Stakeholders' (13), and 'Software debugging' (5), while the rest of the subcategories covered 4 or fewer primary studies.

Fig. 8 shows the number of SRs for each of the sixteen top-level categories. Individual Human Aspects had the largest number of SRs (19) followed by Team with 13 SRs, while Management and Software Development (Workflow) had 10 SRs each, and Success with 9. As for Agile, Knowledge Sharing, and Social and Innovative each domain was covered by 8 SRs. This was followed by Other with 6, and Experience and Productivity with 5 each. Economic Factors and Technical Human Aspects were covered by 3 SRs each, followed by Environmental Factors with 2 SR. The remaining domains were Soft Skills and Software Engineer Controllers covered by one SR each.

4.2. Research question 2

The second research question is: *What is the profile of the most influential and active secondary study researchers in the software engineering human aspects domain?*

This research question seeks the most prominent researchers and organisations in the field, and the most popular research venues. To answer RQ2 with regard to researchers: Three individuals co-authored three SRs each, including Chiranjeev Kumar (Shameem et al., 2018; Sinha et al., 2020; Shameem et al., 2017), Fabio Q.B. da Silva (da Silva et al., 2010; Onoue et al., 2018; Cruz et al., 2011), and Rafael Prikladnicki (Silveira and Prikladnicki, 2019; da Silva et al., 2010; Dutra et al., 2015). Mohammad Shameem authored one SR (Shameem et al., 2018) and co-authored two others (Sinha et al., 2020; Shameem et al., 2017). Sikandar Ali authored two SRs (Ali and Khan, 2014; Ali et al., 2017) and co-authored one (Abrar et al., 2019). Other individuals who co-authored two SRs each included Suriayati Chuprat (Fatima et al., 2019; Nazir et al., 2019), Tracy Hall (Sharp et al., 2009; Beecham et al., 2008), Bibhas Chandra (Shameem et al., 2018, 2017), César França (Dutra et al., 2015; Costa and França, 2019), Emilia Mendes (Vishnubhotla et al., 2018; Britto et al., 2014), Hugh Robinson (Sharp et al., 2009; Beecham et al., 2008), Kai Petersen (Garousi et al., 2016; Jabbari et al., 2016), Lars Göran Wallgren (Lenberg et al., 2015, 2014), Luiz Fernando Capretz (Onoue

Table 10
Top cited SRs on human aspects in software engineering.

SR ID	Citations	Number of authors	Number of institutions	Number of countries
Dybå and Dingsøyr (2008)	3062	2	1	1
da Silva et al. (2010)	1681	4	2	1
Beecham et al. (2008)	602	5	2	1
Pedreira et al. (2015)	382	4	2	1
Walia and Carver (2009)	233	2	2	1
Sharp et al. (2009)	214	5	3	1
Jabbari et al. (2016)	196	4	2	2
Lenberg et al. (2015)	189	3	3	1
Onoue et al. (2018)	177	3	3	2
Garousi et al. (2016)	128	3	4	3

et al., 2018; Jia et al., 2016), Nathan Baddoo (Sharp et al., 2009; Beecham et al., 2008), Robert Feldt (Lenberg et al., 2015, 2014), and Siffat Ullah Khan (Ali and Khan, 2014; Ali et al., 2017). Three individuals authored one SR and co-authored another, including Nargis Fatima (Fatima et al., 2019; Nazir et al., 2019), Sarah Beecham (Sharp et al., 2009; Beecham et al., 2008), and Sumaira Nazir (Fatima et al., 2019; Nazir et al., 2019), while Per Lenberg authored two (Lenberg et al., 2015, 2014). All other authors were involved with one SR each.

Additionally, there were 166 unique individuals who authored and co-authored the 67 SRs, with the largest number of authors on a single SR being eight (Abrar et al., 2019); the smallest being one (Senapathi and Srinivasan, 2013), and the average authors per publication being 3.31. The 67 SRs all together accounted for a total of 6,582 citations (Google Scholar), with an average of 101.26 citations each and median of 18.01. The SR titled "Empirical studies of agile software development: A systematic review" (Dybå and Dingsøyr, 2008) by Dybåand Dingsøyr published in the journal *Information and Software Technology* in 2008 was the single most cited SR with 3,062 citations for approximately 47% of all citations across the 67 SRs. Table 10 summarises the ten most-cited SRs on human aspects in software engineering (refer to Human Aspects Taxonomy in the replication package (Zolduoarrati et al., 2022) for a full comprehensive list of human aspects SRs).

With regard to organisations, the SRs in the study sample emerged from a total of 78 institutions across 27 countries: Australia, Belgium, Brazil, Canada, China, Ecuador, Finland, Germany, Hong Kong, India, Indonesia, Ireland, Japan, Malaysia, Mexico, Morocco, New Zealand, Norway, Pakistan, Romania, Russia, Saudi Arabia, Spain, Sweden, Turkey, the United Kingdom, the United States, and Uruguay. The Federal University of Pernambuco, Brazil led the group with eight SRs across 19 authors, followed by Blekinge Institute of Technology, Sweden, with seven SRs across 10 authors. Four institutes had three SRs each, including Chalmers University in Sweden (3 authors), University of Gothenburg in Sweden (2 authors), University of Malakand in Pakistan (3 authors), and University of Swat in Pakistan (6 authors). Institutes with two SRs each included Federal University of Pará (7 authors), Indian Institute of Technology (7 authors), Mississippi State University (2 authors), Monash University (5 authors), National University of Modern Languages (2 authors), Universiti Teknologi Malaysia (1 author), University of Brasília (2 authors), University of Hertfordshire (3 authors), University of Petroleum (4 authors), University of Rio Grande do Sul (4 authors), Western University (1 author) and University of Alabama (1 author). The remaining 60 institutes participated in one SR each.

Geographically, Brazil had the most SRs (17) coming from 12 of its universities. The popularity and evolution of human aspects research in the Brazilian IT industry has been documented (Herbsleb et al., 2001). The next highest number of SRs (seven) came from Sweden (three universities) and Pakistan (seven universities). Malaysia came next with five SRs from five universities, and Germany with four SRs from three universities and one institute. The following countries had three SRs each, including Australia from two universities, Canada from one institute and one university, China from two universities, India from two universities, Norway from one institute and two universities, and the United States from six universities. The following countries had two SRs each, including Finland from three universities, Mexico from two universities, New Zealand from two universities, Turkey from three universities, and the United Kingdom from three universities. The remaining countries each published one SR (see Fig. 9).

Answering RQ2: Regarding most popular publication venue: In terms of the journals and conferences publishing the SRs, the *Information and Software Technology* journal published the largest number of the SRs at nine (Walia and Carver, 2009; Garousi et al., 2016; Anu et al., 2018; Dybå and Dingsøyr, 2008; Pedreira et al., 2015; Sharp et al., 2009; Beecham et al., 2008; Sánchez-Gordón and Colomo-Palacios, 2019; Heaton and Carver, 2015), followed by *IEEE Access* with four (Anchundia and Fonseca, 2020; Abrar et al., 2019; Ali et al., 2017; Anwar et al., 2019), followed by three SRs each from the *International Conference on Global Software Engineering* (da Silva et al., 2010; Ali and Khan, 2014; Kroll et al., 2013), the *Brazilian Symposium on Software Engineering (SBES)* (Veras et al., 2020; Flores and de Alencar, 2020; Canedo and Santos, 2019), the *Euromicro Conference on Software Engineering and Advanced Applications* (Cruz et al., 2011; Ciupe et al., 2017; Chagas et al., 2014), and *IEEE Transactions on Software Engineering* (Dutra et al., 2015; Melo et al., 2012; Lima et al., 2019). The *International Conference in Software Engineering Research and Innovation (CONISOFT)* (Ramírez-Mora and Oktaba, 2017; Durán et al., 2019) and the *International Conference on Engineering Technologies and Applied Sciences (ICETAS)* (Fatima et al., 2019; Nazir et al., 2019) each published two SRs. The remaining venues published one SR each. Furthermore, we have also plotted a heatmap as a visualisation to portray which countries are more concerned with which human aspects, (refer to Heatmap in the replication package (Zolduoarrati et al., 2022) for the rendered heatmap).

4.3. Research question 3

The third research question is: *What are the quality levels of the secondary studies conducted in the software engineering human aspects domain?*

The quality scores of the individual SRs included in this review were presented in Fig. 2. Answering RQ3: Table 11 below answers RQ3, where it summarises the average quality scores of all the SRs included in this study by publication venue.

Conferences, journals, theses, and workshops are each fundamentally different (e.g., in terms of scope and size), and therefore we have classified them separately before quantifying the average quality score. Two approaches were taken to investigate the quality levels of the SRs. The first approach measured the mean quality score for each venue type, as done by Hoda et al.

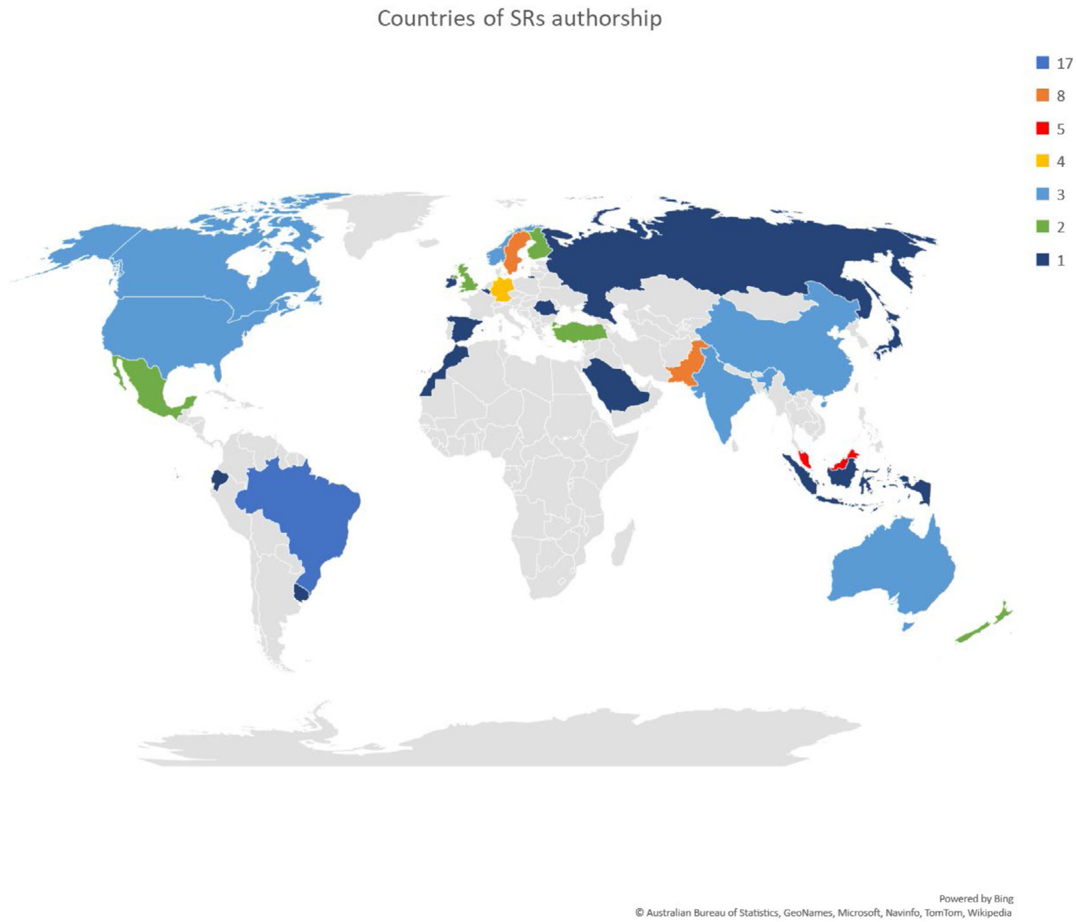

Fig. 9. SRs publications by country.

Table 11
SRs count and mean quality per publication venue.

Publication venue	Count	Mean
Conference	38	3.01
Journal	25	3.22
Thesis	1	3.54
Workshops	3	2.67
Overall	*67*	*3.09*

(2017), who quantified quality scores for 28 SRs on agile software development. The mean quality scores across venue types were different. The highest average quality score from the secondary studies was 3.54, which was found in the thesis, because the thesis fulfils most of the quality assessments from Q1 to Q4 due to the study typically being performed over a larger timeframe, having broader scope, and having more space to report details. Journal papers had an average quality score of 3.22, followed by conference (3.01) and workshop papers (2.67). An overwhelming 88.1% of all SRs were conducted using the guidelines from Kitchenham (2004), Kitchenham and Charters (2007a), Centre for Reviews and Dissemination (2016), Beecham et al. (2008), Petersen et al. (2015), Dyba et al. (2007), Webster and Watson (2002), or Brereton et al. (2007), leading to a significant contribution in standardising and improving the guidelines for secondary studies and quality in empirical studies.

Reflecting on the work by Hoda et al. (2017) and the guidelines provided by Kitchenham and Charters (2007a), the second approach investigated Pearson's correlation between quality score and different categories of study type (empirical, theoretical, experimental, or other study types), subject of investigation (industry, non-industry, both, or other subject of investigation), and methods (experiment, case study, reports, ethnography, survey, literature review, solution proposal, dataset, interview, observation, multiple methods approach, or other methods). The null hypothesis was that the correlation was zero and the alternate hypothesis was that the correlation was not zero. The decision was made based on a 5% significance level. With regards to study type, no statistically significant correlation was found between three of the study types (empirical, theoretical and experimental) and quality score, either in conference, journal, thesis, or workshops. With regards to subject of investigation, the only statistically significant correlation found was between "Both" (industry and non-industry) and the quality score in journal (0.452) and conference (−0.374) papers. The relationship with conference papers was negative, meaning that the secondary studies focussing on both industry and non-industry as the subject of investigation are statistically associated with lower quality score. Finally, in terms of method, the only significant correlation found was between multiple methods approach and quality score in conference papers (0.483). This finding means that studies adopting a multiple methods approach are statistically associated with higher quality score.

In conclusion, based on the total secondary studies found across all venue types, empirical study was the most common study type used (1,129 studies), followed by other study types (303 studies), theoretical study (140 studies) and experimental study (14 studies). Furthermore, in terms of subject of investigation, industry was the most frequent focus of investigation across all paper types (505), followed by other subject of investigation (503), non-industry (385), and both (204). Finally, in terms of method, multiple methods approach was the most frequently adopted method in secondary studies (918), followed by case study (447) and questionnaire (420). The quality scores of papers across publication venues were competitive. The highest average of quality score from the secondary studies was found in the thesis venue type, followed by journal, conference and workshops. Overall, there was no systematic pattern regarding the correlation between study types (empirical, theoretical, experimental and other study types) and quality score, between the subject of investigation (industry, non-industry, both, or other subject of investigation) and quality score, or between method (experiment, case study, reports, ethnography, survey, literature review, solution proposal, dataset, interview, observation, other methods, and multiple methods approach) and quality score.

4.4. Research question 4

The fourth research question is: How are various dimensions of human aspects studies related, and what predicts study quality?

This question has a two-fold purpose. First, it aims to investigate the relationship among the attributes of human aspects studies, which are total quality score, different categories of study types (empirical, theoretical, experimental, or other study types), subjects of investigation (industry, non-industry, both, or other subject of investigation), and methods (experiment, case study, reports, ethnography, survey, literature review, solution proposal, dataset, interview, observation, multiple methods approach, or other methods) As seen in Table 2, all these variables were utilised to conduct the required analysis for RQ4. Furthermore, these variables are sourced from the underlying secondary studies. Second, RQ4 aims to investigate the determinants of study quality.

For the first objective, Pearson correlation analysis was applied. Pearson correlation analysis is used to understand the linear relationship among a high number of variables (Schober et al., 2018). The null hypothesis (H0) is that the correlation between two variables is statistically not different from zero. The alternate hypothesis (H1) is that the correlation between two variables is statistically different from zero (Schober et al., 2018). Pearson correlation is appropriate when all the variables in the examined correlation matrix are ratio scale. A statistically significant correlation score is assessed in terms of its magnitude. In this study, the magnitude classification follows Cohen (1988) where the correlation is considered low if the value is less than 0.30 ($0 < r < 0.30$); moderate if the correlation is between 0.30 and 0.50 ($0.30 \leq r < 0.50$); and strong if the correlation value is higher than or equal to 0.50 ($0.50 \leq r < 1.00$).

For the second objective, a multiple linear regression was estimated using ordinary least squares (OLS) to measure the linear relationship between quality score and other attributes. Positive and negative trends observed from plotted data indicated a possibility for linear approximation to investigate the attributes of a study's quality score. Multiple linear regression is widely used to investigate the determinants of a particular variable (Slinker and Glantz, 2008) due to its applicability. There are other higher-level approximations that can be used, such as polynomial regression, but linear regression is preferred in this study due to its simplicity and applicability. Also, the dependent variable in the regression (the quality total score) is not a binary variable, making Tobit, Probit, or logistic regression less preferred (Slinker and Glantz, 2008; Glantz and Slinker, 2001). The multiple linear regression model is constructed from the information obtained from SRs using quality score as the dependent variable. Independent variables included number of primary studies covered, number of years covered, number of authors, total citations, total institutions, total countries, total number of topics, total study types, total subjects of investigation, and total methods.

In addition to independent variables, the model also incorporates one categorical variable of publication venue type, which consists of journal, conference, workshop, and thesis. These four venue types are coded as dummy variables and generated from a single variable (i.e., venue type). However, the venue type "thesis" was omitted from the equation because it functions as a reference for journal, conference, and workshop. Finally, we maintain the linear time effect in the regression by incorporating publication year as one of the regressors in the model (Diebold, 1998). Some estimation techniques are also applied to satisfy OLS assumptions[2] and ensure that OLS is the best linear unbiased estimator.

However, it is also important to note that due to the small number of observations, the model does not incorporate the detailed topics, detailed study types, detailed subject of investigation, or detailed methods as regressors. Only the totals from all these variables are incorporated as regressors. This is because these variables themselves consist of 35 variables in total, which can take more than half of the degrees of freedom if incorporated, and thus significantly reduce the model's predictive ability. Moreover, incorporating all these items can overfit the model, destabilise the estimated coefficients, and thereby further reduce the model's predictive ability (Hawkins, 2004; Gujarati, 2003). Table 12 provides summary statistics for the variables utilised in the study.

Answering RQ4: In terms of relations between variables: The detailed results of our correlation analysis can be seen in Table 13. Additionally, the full correlation matrix may be seen in Calculations in the replication package (Zolduoarrati et al., 2022). Based on the correlation scores, there were 51 pairs having strong correlation (i.e., $r \geq 0.5$). Among these, some interesting pairs stand out. Correlation analysis shows that among subjects of investigation, industry has greater correlation with some study types compared to non-industry. Specifically, the correlation score between industry and empirical is 0.652, between industry and theoretical is 0.71, and between industry and experimental is 0.749. In contrast, the correlation score between non-industry and empirical is 0.338, between non-industry and experimental is 0.384, and between non-industry and theoretical is less than 0.30. SRs that investigated both industry and non-industry are strongly correlated with empirical (0.56) and experimental (0.538) study types, although the magnitudes are weaker than for SRs that investigated industry alone.

The correlation analysis also found that Individual Human Aspects is the only topic that strongly correlated with both empirical (0.539) and experimental (0.55) study types. Besides Individual Human Aspects, the analysis also found that the Environmental Factors topic is strongly correlated with the workshops publication venue (0.566).

With regards to the medium correlation group, some interesting findings also stand out. First, different from the strong correlation group, empirical and experimental study types are positively correlated with non-industry (0.338 and 0.384, respectively), while industry is only correlated positively with thesis (0.335).

[2] The OLS assumptions are as follows: First, the variance inflation factor (VIF) of each independent variable was measured to avoid multicollinearity issues (Diebold, 1998). Second, outliers were removed based on studentized residuals criteria where any observations with studentized residuals higher than 2 are omitted (Thompson et al., 2017; Paul and Fung, 1991). Third, the Breusch–Pagan test was applied to avoid any heteroskedasticity issue (Osborne and Waters, 2002). Fourth, Skewness and Kurtosis normality tests were applied to ensure that residuals of the model are normally distributed (Breusch and Pagan, 1979).

Table 12
Statistical summary.

	Mean	Interpretation
Quality total score	3.09	the mean quality score was 3.09 on a scale of 0 to 4
Primary studies covered	66.06	the average number of primary studies covered was 66 studies per SR
Total years	16.06	the number of years covered on average was 16 per SR
Authors	3.31	one SR on average contains 3 authors
Citations	101.26	one SR on average contains 101 citations
Institutions	1.69	one SR on average contains 2 institutions
Countries	1.22	one SR on average contains 1 country
Total topics	184.85	one SR on average contains 184 topics
Total study types	29.76	one SR on average contains 30 study types
Total subjects investigated	16.33	one SR on average contains 16 subjects of investigation
Total methods	34.52	one SR on average contains 34 methods

We avoid retaining all the venue types simultaneously in the model as this causes intercept drop-off and prevents the application of the Breusch–Pagan heteroskedasticity test, which is a crucial test of the OLS assumptions. The results of the OLS regression are shown in Table 14. The model performance is 0.512 or 51.2%, meaning the independent variables in the model can explain 51.2% of variation of quality total score. The F-statistic also has a significant p-value < 0.01, indicating the linear combination of regressors and dummy variables in the regression are significant in affecting the dependent variable. The model has satisfied all OLS assumptions, meaning all regressors involved in the model are free from multicollinearity issues. The residuals generated from the model are also homoskedastic, as the Breusch–Pagan heteroskedasticity test delivered a p-value higher than the 1%, 5%, and 10% significance levels, confirming the homoskedastic null hypothesis. Finally, the skewness and kurtosis test, which aims to investigate the normality distribution of the residuals, also delivered a p-value higher than any significance level in the OLS model, confirming the null hypothesis that the residuals are normally distributed.

In terms of variables predicting study quality: Some of the variables impact quality score positively while others impact it negatively as seen by the sign of the coefficients in Table 14. In terms of positive impact, the model demonstrates that number of countries and total topics covered significantly correspond to higher quality total score. Specifically, the unstandardised coefficient[3] for countries demonstrates that a one-country increase in the secondary study significantly corresponds to an increase of the paper quality score by 0.440, everything else being equal. Furthermore, the number of total topics is also positively significant in affecting quality total score. One additional topic included in an SR significantly corresponds to an increase of quality total score by 0.001, everything else being equal.

Some negative determinants of the quality total score are shown in primary studies covered, total years covered, and the number of institutions. Specifically, the unstandardised coefficients show that an increase of one primary study significantly reduces the quality total score by 0.00441, everything else being equal. Similarly, a one-year increase of total years covered significantly reduces the quality total score by 0.0217, everything else being equal. Furthermore, one additional institution significantly reduces the quality total score by 0.242, everything else being equal. An important note is required to explain the relationship between the number of institutions and the number of countries. Although the two are positively correlated, the negative impact from having one more institution on the quality score of a secondary study happens when the number of countries is fixed. This is important because more countries are associated with more institutions. Thus, the opposite sign of impact between number of countries and number of institutions on the quality total score is surprising given that the relationship between number of countries and number of institutions is positive, while the impact of number of countries on quality total score happens only when the number of institutions is unchanged, and vice versa.

Based on the unstandardised coefficients, it seems that more primary studies, more institutions involved, and more years covered do not necessarily increase the total quality score. However, based on the standardised coefficients, fewer total years covered is more effective than reducing the number of institutions or primary studies covered. Specifically, reducing the primary studies covered, total years covered, and number of institutions by one standard deviation increases the total quality score by 0.464, 0.366 and 0.364 respectively, everything else being equal. This finding implies that reducing the number of primary studies covered is 1.26 times or 26% more effective than reducing the total years covered. Similarly, reducing the number of primary studies covered is 1.27 times or 27% more effective than reducing the number of institutions. Additionally, the OLS regression model demonstrated that publication year is negative and statistically significant, where one year forward (e.g., one year more recent) lowers the quality score by 0.0528 on average, indicating that the average quality score tended to diminish in recent studies.

5. Limitations

This tertiary review analysed secondary studies relating to human aspects in software engineering. Regarding study selection validity, we have iteratively evaluated our search string to yield fine grained results as seen in Table 3, before eventually running the search string on credible digital libraries such as ACM Digital Library, IEEE Xplore, and ScienceDirect (refer to Search within the replication package for details of each database search queries and retrieved papers) without year restrictions to ensure the validity of the results. Such digital libraries were also employed in other SRs (e.g., Kitchenham et al., 2010). One of the primary search expressions adopted in this investigation was 'human aspects', which was then incorporated with other terms and morphemes such as 'human factors'. Despite utilising this strategy, several SRs on human aspects were still omitted during the initial examination. Such papers did not include 'human aspect' in the title, abstract, or keywords. Thus, we have performed backwards snowballing (as recommended by Ampatzoglou et al., 2020) to include omitted publications. Still, various "human" terminology was used in secondary publications, making investigators regularly question whether such terminologies represent the same concepts. This additional interpretative overlay produces further chances for inaccuracies within the categorisation procedure. Also, not employing a detailed search mechanism and string to gather a large sample of evidence may have overlooked related publications in different scientific outlets, since relevant publications may not be listed in the databases searched. Restricting our search to 'computer science' or related domains (see

[3] Unstandardised coefficients are those obtained after running a regression model on variables measured in their original scales, as opposed to being rescaled and standardised to a mean of 0 and a standard deviation of 1.

Table 13
Correlation matrix for strong and moderate variables.

Strong correlations			Moderate correlations		
Both industries	Total subject of investigation	0.670	Institutes	Journal	0.396
Both industries	Empirical	0.506	Both industries	Journal	0.452
Both industries	Experimental	0.538	Both industries	Institutes	0.342
Both industries	Industry	0.594	Both industries	Countries	0.310
Case study	Total Methods	0.773	Both industries	Total study types	0.390
Conference	Journal	−0.881	Both industries	Individual Human Aspects	0.391
Countries	Institutes	0.583	Both industries	Theoretical	0.490
Empirical	Total study types	0.712	Both industries	Non Industry	0.461
Empirical	Total subject of investigation	0.629	Both industries	Conference	−0.374
Empirical	Individual Human Aspects	0.539	Case study	Primary studies covered	0.441
Environmental Factors	Workshops	0.566	Case study	Other	0.396
Ethnography	Total Methods	0.687	Case study	Experiment	0.452
Ethnography	Other	0.513	Countries	Total Years	0.315
Ethnography	Case study	0.818	Countries	Authors	0.325
Experimental	Total study types	0.651	Economic Factors	Thesis	0.455
Experimental	Total subject of investigation	0.772	Empirical	Journal	0.368
Experimental	Individual Human Aspects	0.550	Empirical	Conference	−0.392
Experimental	Empirical	0.739	Ethnography	Primary studies covered	0.479
Experimental	Theoretical	0.878	Ethnography	Experiment	0.467
Industry	Total study types	0.501	Experience	Total Years	0.306
Industry	Total subject of investigation	0.902	Experiment	Total Methods	0.433
Industry	Individual Human Aspects	0.509	Experimental	Primary studies covered	0.332
Industry	Empirical	0.652	Individual Human Aspects	Conference	−0.304
Industry	Theoretical	0.710	Individual Human Aspects	Primary studies covered	0.320
Industry	Experimental	0.749	Individual Human Aspects	Total study types	0.386
Interview	Success Factors	0.553	Individual Human Aspects	Total subject of investigation	0.491
Literature review	Total Methods	0.516	Industry	Thesis	0.335
Literature review	Case study	0.647	Institutes	Conference	−0.365
Management	Thesis	0.624	Institutes	Total Years	0.357
Multiple methods approach	Total subject of investigation	0.633	Institutes	Authors	0.439
Multiple methods approach	Agile	0.614	Interview	Total Topics	0.439
Non-Industry	Total subject of investigation	0.690	Interview	Literature review	0.484
Observation	Dataset	0.975	Interview	Dataset	0.387
Other methods	Total Methods	0.707	Literature review	Total Topics	0.388
Other methods	Social and Innovative	0.565	Literature review	Success Factors	0.422
Other methods	Other study types	0.857	Literature review	Ethnography	0.453
Other study types	Total study types	0.562	Literature review	Questionnaire survey	0.331
Other study types	Social and Innovative	0.673	Management	Total Topics	0.355
Other Topics	Primary studies covered	0.539	Multiple methods approach	Total study types	0.422
Questionnaire/survey	Total Methods	0.519	Multiple methods approach	Empirical	0.322
Software Development (Workflow)	Thesis	0.914	Multiple methods approach	Theoretical	0.440
Software Development (Workflow)	Management	0.605	Multiple methods approach	Industry	0.421
Solution proposal	Thesis	0.867	Multiple methods approach	Both industries	0.317
Solution proposal	Management	0.523	Multiple methods approach	Conference	0.483
Solution proposal	Software Development (Workflow)	0.803	Non Industry	Empirical	0.338
Success Factors	Total Topics	0.739	Non Industry	Experimental	0.384
Theoretical	Total study types	0.677	Non Industry	Industry	0.433
Theoretical	Total subject of investigation	0.681	Observation	Interview	0.395
Theoretical	Individual Human Aspects	0.555	Other topics	Experience	0.372
Theoretical	Empirical	0.777	Other methods	Primary studies covered	0.368
Total subjects of investigation	Total study types	0.528	Other methods	Total Topics	0.326
			Other methods	Total study types	0.454
			Other methods	Ethnography	0.364
			Other methods	Questionnaire survey	0.426
			Other study types	Total Topics	0.357
			Other study types	Total Methods	0.403
			Primary studies covered	Conference	−0.310
			Questionnaire survey	Total Topics	0.330
			Questionnaire survey	Team	0.331
			Questionnaire survey	Individual Human Aspects	0.331
			Questionnaire survey	Experiment	0.308
			Questionnaire survey	Case study	0.446
			Reports	Authors	0.414
			Reports	Management	0.401
			Social and Innovative	Total study types	0.404
			Software Development (Workflow)	Total Topics	0.341
			Software Development (Workflow)	Economic Factors	0.396

(*continued on next page*)

Table 13 (*continued*).

	Solution proposal	Economic Factors	0.379
	Solution proposal	Industry	0.303
	Success Factors	Total Methods	0.304
	Technical Human Aspects	Institutes	0.354
	Theoretical	Management	0.316
	Total Methods	Primary studies covered	0.490
	Total Methods	Total Topics	0.450
	Total study types	Total Topics	0.379
	Total Topics	Thesis	0.364
	Total Topics	Primary studies covered	0.324

Table 14
OLS regression results.

Variables	Unstandardised coefficients		Standardised coefficients		Sig.
	Beta	Std. Error	Beta	t stat	
Constant	109.300	50.700		2.156	0.036
Publication year	−0.052	0.025	−0.284	−2.093	0.042
Primary studies covered	−0.004	0.001	−0.464	−3.333	0.001
Total years	−0.021	0.007	−0.366	−3.056	0.003
Authors	0.025	0.064	0.047	0.394	0.696
Citations	0.000	0.000	0.008	0.063	0.949
Institutes	−0.242	0.105	−0.364	−2.313	0.025
Countries	0.440	0.163	0.374	2.695	0.009
Total topics	0.001	0.000	0.358	2.622	0.012
Total study types	0.000	0.002	0.029	0.186	0.853
Total subject investigation	0.000	0.002	0.047	0.311	0.757
Total methods	0.000	0.001	0.078	0.556	0.581
Paper type					
Journal	0.651	0.606	0.509	1.045	0.302
Conference	0.306	0.723	0.244	0.505	0.616
Workshops	0.677	50.7	0.195	0.937	0.354
Observations	60[a]	F-stat	3.380		
R-squared	0.512	Sig >F	0.000		

[a]Seven observations are omitted due to studentized outliers.

Table 4) may have omitted related publications in other domains. We adopted several strategies to enhance external validity, such as examining the profiles of the authors for any other related publications, being adequately versed with human aspects topics, and discussing extracted studies as a group. These mitigations were done to include as many SRs associated with human aspects as possible, in turn ensuring comprehensive coverage and generalisable high-quality conclusions based on the large sample. However, despite these interventions, search results from the digital libraries were evaluated non-empirically, and as such, results from each iteration of the search string refinement were not fully documented.

Aside from search string validity, however, there are still several limitations. For instance, as tertiary studies regarding human aspects in software engineering are scarce, this study did not incorporate a ‘gold-standard’ to compare our results to, as recommended by Ampatzoglou et al. (2020). Moreover, due to the researchers’ linguistic background, only studies written in English are included in this study. In addition, as stated in Section 3.3, we may have missed a small number of relevant studies due to certain papers being outside our university’s subscription. Nevertheless, the studies missed are not expected to be numerous, given our coverage of all the top databases. Finally, some related publications may be subject to publication bias since grey literature (e.g., World Bank annual reports) were excluded, due to their nature being non-compliant with the goals of the study.

Regarding data validity, our inclusion/exclusion criteria in Table 4 were enhanced by backwards snowballing to achieve the highest possible sensitivity and precision, as discussed by Cruz et al. (2015). While initially we had differing opinions regarding inclusion/exclusion criteria, we discussed those processes and resolved conflicts or disagreements in joint meetings. During the quality assessment process, disagreements between raters were measured by Cohen’s Kappa (refer to Quality in the replication package (Zolduoarrati et al., 2022) for details on Cohen’s Kappa calculation). Subsequently, the categorisation of the research domains (RQ1) was conducted by identifying the research topics included throughout different secondary studies and reviewing them amongst the authors, followed by grouping similar topics collectively below the corresponding top-level category, which was identified from the overall research focus of the SR. However, the categorisation process could be interpreted differently by others. Some categories might be perceived as less dominant than others despite the majority of the SRs being clearly focused on a single research area. To mitigate this possible issue, the authors discussed the categorisation process and shared suggestions in multiple review meetings where no notable conflicts in the categorisation procedure were noted.

Regarding the validity of the SRs, some did not adequately report or justify the original reasons for selecting the specific aspects covered, their relationship with other similar aspects and how they came to existence, or how they contributed to the software engineering discipline, thereby causing a threat to internal validity. We have not been able to control this threat independently since the authors of these secondary studies might not report the original reasons for a variety of reasons, such as a relationship with a particular industrial vendor.

For RQ3, we have employed mean values of the quality scores as is standard practice (Kitchenham et al., 2010). However, doing so implies a ratio (interval), while the quality is considered as an ordinal measure. Thus, quality rating enabled us to rank publications based on lower or higher quality and compare them to each other.

It is also crucial to understand that the conclusions of this study should be viewed in light of human aspects being a complex phenomenon. Interpreting it through labels such as Success

Factors, Productivity, and Individual Human Aspects is insufficient and presents limited knowledge regarding the mechanisms and causes underlying these phenomena (Meldrum et al., 2017). This is particularly true when most human aspects have been studied in complete isolation from each other (Mohanani et al., 2018). Human aspects SRs do not present a comprehensive theory of reasoning behind the selected primary studies. They are empirical generalisations regarding common problems or topics and, despite such problems or topics seeming to be related, their relationships and causes are vague. In short, synthesising research based on human aspects may be assessed as epistemologically fraught.

In conclusion, future works may draw inspiration from this manuscript as we have provided details in our analysis, starting from research question formulation up until statistical inference. Furthermore, more than one researcher was involved in the analysis and inference. We have potentially mitigated researcher bias and have provided results as objectively as possible. As such, most of the work we have done is repeatable while only certain parts are not fully documented (e.g., search string evaluation). We can therefore say that the answers to this study's research questions lead to the achievement of the work's overall goal as discussed in Section 1.

6. Conclusion

We conducted a tertiary study of secondary studies (SRs) related to human aspects in software engineering using the guidelines put forth by Kitchenham et al. (2010). RQ1 asked *what commitment has been invested to gather evidence in software engineering human aspects and the discussed domains?* Through this tertiary review, 67 secondary studies were identified that investigated various human aspects. As a result, 16 distinct top-level categories were identified in the literature: *agile, economic factors, environmental factors, experience, individual human aspects, knowledge sharing, management, productivity, social and innovation, soft skills, software development (workflow), software engineer controllers, success factors, team, technical human aspects*, and *other*. Of these 16 categories, *individual human aspects* was investigated most often (17.4% of the SRs). Other key topics investigated in the SRs were *team* (11.9%), *management* (11%), and *software development* (*workflow*) (11%). Rarely discussed domains were *soft skills* and *software engineer controllers* (both appearing in 0.9% of SRs). These trends highlight the topics that have received the most research to date and the areas that could use further attention.

RQ2 asked *what is the profile of the most influential and active secondary study researchers in the software engineering human aspects domain?* We found that Mohammad Shameem, Sikandar Ali, Per Lenberg, and Shirley Cruz were the most prolific SR authors, with two publications each. Chiranjeev Kumar, Jeffrey C. Carver and Rafael Prikladnicki co-authored the highest number of SRs (three). The Federal University of Pernambuco, Brazil was involved in the most SRs (eight), followed by Blekinge Institute of Technology, Sweden, with seven. The SR carried out by Dybå and Dingsøyr (2008) titled "Empirical studies of agile software development: A systematic review" and published in 2008 in *Information and Software Technology* was the single most cited SR. It had 3,062 citations, accounting for nearly 47% of all citations across the 67 SRs.

RQ3 asked *what are the quality levels of the secondary studies conducted in the software engineering human aspects domain?* We found that the average quality score across the investigated SRs was 3.09 out of 4.0, with journal-published SRs having somewhat higher quality (3.22) than conference papers (3.01). In contrast, workshops had the lowest average quality (2.67), while the sole thesis had the highest quality (3.54). The SRs were conducted using the Kitchenham guidelines (Glasziou et al., 2000; Kitchenham and Charters, 2007a; Kitchenham et al., 2009), followed by Beecham et al. (2008), Petersen et al. (2015), Dyba et al. (2007), Webster and Watson (2002) and Brereton et al. (2007), which have all been effective in systematically gathering evidence on human aspects in software engineering.

RQ4 asked *how are various dimensions of human aspects studies related, and what predicts study quality?* We discovered that the relationships among variables were diverse. Some of them were positive while others were negative, but most of the correlation scores were weak. Using Cohen's classification (Cohen, 1988), strong correlation scores were found between many pairs of variables, such as between Journal and Conference; Software Development (Workflow) and Thesis; Solution Proposal and Thesis; Total Study Types and Empirical; Total Study Types and Theoretical; Industry and the Total Subject of Investigation; Other Methods and Other Studies; Both and Experimental; Both and Industry; Ethnography and Case Study; and Observation and Dataset, among other correlations.

Because interest in the human aspects related to software engineering has risen significantly in recent years, we suggest that further research be done on human aspects in the less investigated topics (e.g., Economic Factors, Soft Skills, Environmental Factors, and Productivity). These topics are highly relevant to software engineering practices and are excellent candidates for in-depth research to better understand how these topics might improve the state of the field. Additionally, because none of the SRs revealed a focus on multiple human aspects (instead focusing primarily on a single human aspect), research into collections of human aspects and how they interact and might influence software engineering would be a worthwhile pursuit.

6.1. Implications

The findings of this paper have a number of implications for both research and practice. In terms of research, this paper has identified the current state of knowledge on human aspects in the software engineering domain. This has allowed for a good understanding of current research trends, literature gaps, and the potential for future research. Furthermore, it may help scholars in the field to easily identify the trending research domains, thus allowing them to more effectively build on this research and deepen their understanding of the field. Our work has also revealed research that will be helpful in practice. This tertiary review aggregates research related to human aspects in software engineering, which might help practitioners quickly identify relevant research. Because a growing body of research has shown the important role of human aspects in software engineering (Guveyi et al., 2020; Abreu and Premraj, 2009; Al-Rawas and Easterbrook, 1996), learning more about these factors through a broad overview of the field might help practitioners discover new ways to enhance the quality of software engineering projects.

6.1.1. Implications for software engineering research

This study is, to the best of the authors' knowledge, the only wide-ranging tertiary study to date on human aspects in the software engineering field. Thus, it has significant implications for software engineering research in terms of identifying research gaps, showing how the discipline has progressed, revealing best practices for SRs in this field, and helping researchers better understand the state of the field and the quality of the research that has been done. It offers a broad view of human aspects and how they affect software practices, thus allowing for a better understanding of human aspects and how they might enhance the quality of the software engineering field. This study also took an approach that allowed for a wider body of scholarly work to be examined, including published research in journals, conferences, workshops, and theses. This allows for important research to be included that might otherwise be overlooked.

One major gap that has been identified in this review that warrants further research is related to terminology. This study began with the intent of understanding "human aspects" in software engineering. However, preliminary searches revealed a variety of search terms that were equivalent to the original query. These included "human aspects", "human issues", and "human errors". However, one limitation is the nature of these terms and whether they truly have identical meanings, and as such we have opted to adopt the definition from Guveyi et al. (2020) Fernández-Sanz and Misra (2011). We propose future research to determine whether there may be other similar terms, what differences there are between them, and how they are used in scholarship. This will ensure that related and relevant studies are not overlooked and provide a potential grouping of studies that will benefit researchers looking for more specific literature.

This review has also revealed that certain human aspects have been studied in isolation. We therefore recommend that future research investigate combinations of human aspects to help identify potential relationships among them.

A number of the key findings from this review will benefit future researchers. First, the broad overview this review provides will save researchers time in familiarising themselves with the state of scholarship on human aspects in software engineering. Second, researchers can gain a better understanding of the current research trends in human aspects in software engineering. This has the potential to help scholars identify the topics that would benefit from further research. Third, this study has identified key scholars and groups of researchers in this field. This will aid in connecting those pursuing similar research topics and help those interested in this research domain quickly identify the best starting points for research (e.g., most-cited studies or most-published authors). Finally, through our quality assessment measures (see Table 5), this review helps researchers to identify high-quality research and to learn how to conduct their own research in a way that achieves the highest quality. For example, SRs with authors from multiple countries tended to be of higher quality. However, given that the number of countries is often proportional to the number of authors, then it is also fair to say that SRs with more authors may tend towards higher quality scores due to the diverse knowledge pool that drives the production wheel. Also, this study found that the number of primary studies covered within an SR does not necessarily correspond to better quality. In other words, the variety of topics is more important than the quantity of primary studies covered, in order to determine the SR's quality score. Also, publication in a journal tended to predict higher quality. Conversely, factors that contributed to lower quality scores included covering too many years or focussing on specific topics rather than a broad view. These findings suggest that quality research uses diverse sources, has a clear, selective scope, and does not try to cover too wide a body of scholarship.

A final trend revealed by this study is that interest in human aspects in software engineering steadily increased from 2007 to 2019. The number of SRs on human aspects in software engineering peaked in 2019 with 12 publications. Thereafter, 2020 saw eight publications, and, to date, there has only been one SR in 2021. The reduced number of publications in 2020 may be a consequence of the COVID-19 pandemic, as researchers may not have been able to carry out as many studies as in the preceding years. Such studies require collaboration, but social interactions during that period were greatly restricted. The meaning of this decline is a topic worth pursuing. In addition to COVID-19, it could be that terminology has changed, which means that further search terms should be identified. It may also mean that research has moved beyond the SR, becoming more nuanced and delving into individual topics highlighted as gaps by the broad overviews presented by the SRs. Other possibilities are that other research areas are trending in software engineering, which suggests a need for future SRs on recent research on human aspects in software engineering.

Because SRs are limited by search terms, available sources, and search date, this study can be replicated in the future. Future studies may also include the results of this tertiary study, particularly related to the trending domains and the quality assessment guidelines.

6.1.2. Implications for software engineering practice

This tertiary review provides the industry with an overview of research on human aspects in software engineering that can provide a starting point for understanding how to use human aspects to improve the industry. It saves the industry time by providing a curated list of quality SRs on human aspects that may help them easily find the information needed to improve human aspects. Practitioners may not have sufficient time to build an effective search query, track down the relevant articles, read them, and form their own conclusions based on the evidence they find in these articles. A tertiary review, therefore, effectively synthesises this information and makes the available evidence accessible to those in the industry.

Another key takeaway for the software engineering industry is that human aspects are widespread and diverse. They likely affect every aspect of software engineering. Thus, those in the industry need to consider human aspects and their effect on software engineering. This tertiary study can help them to accomplish this by providing a broad overview of SRs on human aspects in software engineering. Results from this study may be utilised to guide companies to the literature that may be reviewed and help them to find the information they need quickly to improve or build on the human aspects they intend to pursue. Because human aspects are embedded into all facets of software engineering, the industry may find it useful to devote time to understanding human aspects in order to increase their success.

Additionally, the quality assessment provided in this study can help the industry determine the best research to examine to understand the issues under consideration. By providing practitioners with a body of scholarship and assessing the quality of the included articles, software engineering practitioners can save valuable time in identifying relevant sources.

This tertiary study also highlights the strongest and potentially most useful SRs for practitioners to use when beginning to explore the field. For example, Table 10 lists the most frequently cited SRs, which are likely the most useful for understanding the research in the particular category they cover. These statistics, along with the list of primary researchers and research institutions, can help practitioners to quickly identify the best SRs to use in their preliminary considerations. Average SR quality remained relatively steady from 2007 to 2021, with 2019 having the lowest average quality. This anomaly may be due to certain external factors, prompting the need for a future study to further investigate the quality of the primary studies included in those secondary studies to determine the causes this decline in SR quality.

Overall, our work provides a clear and comprehensive overview of the evidence available on human aspects of software engineering for the community to better understand the current field of research.

Based on this tertiary review, we recommend a stricter quality control in tertiary studies by dismissing SRs that do not follow literature review guidelines or reference their employed primary literature. Using empirical evidence, this tertiary study provided a broad lens examining secondary literature on human aspects in software engineering in academic and industrial settings.

CRediT authorship contribution statement

Elijah Zolduoarrati: Conceptualization, Data curation, Methodology, Writing – original draft. **Sherlock A. Licorish:** Conceptualization, Data curation, Supervision, Writing – review & editing, Statistics validation. **Nigel Stanger:** Data curation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data shared as a replication package online via the Zenodo portal Zolduoarrati et al. (2022).

References

Abrahamsson, P., Warsta, J., Siponen, M.T., Ronkainen, J., 2003. New directions on agile methods: A comparative analysis. In: Proceedings of the 25th International Conference on Software Engineering, 2003. IEEE, pp. 244–254. http://dx.doi.org/10.1109/ICSE.2003.1201204.

Abreu, R., Premraj, R., 2009. How developer communication frequency relates to bug introducing changes. In: Proceedings of the Joint International and Annual ERCIM Workshops on Principles of Software Evolution (IWPSE) and Software Evolution (Evol) Workshops. pp. 153–158. http://dx.doi.org/10.1145/1595808.1595835.

Acuña, S.T., Gómez, M., Juristo, N., 2009. How do personality, team processes and task characteristics relate to job satisfaction and software quality? Inf. Softw. Technol. 51 (3), 627–639. http://dx.doi.org/10.1016/j.infsof.2008.08.006.

Al-Rawas, A., Easterbrook, S., 1996. Communication problems in requirements engineering: A field study. In: Proceedings of the First Westminster Conference on Professional Awareness in Software Engineering (No. NAS 1.26: 203071). Retrieved from https://ntrs.nasa.gov/api/citations/19970005658/downloads/19970005658.pdf.

Ampatzoglou, A., Bibi, S., Avgeriou, P., Chatzigeorgiou, A., 2020. Guidelines for managing threats to validity of secondary studies in software engineering. In: Contemporary Empirical Methods in Software Engineering. Springer, Cham, pp. 415–441. http://dx.doi.org/10.1007/978-3-030-32489-6_15.

Bannerman, P.L., 2008. Risk and risk management in software projects: A reassessment. J. Syst. Softw. 81 (12), 2118–2133. http://dx.doi.org/10.1016/j.jss.2008.03.059.

Beecham, S., Baddoo, N., Hall, T., Robinson, H., Sharp, H., 2008. Motivation in software engineering: A systematic literature review. Inf. Softw. Technol. 50 (9–10), 860–878. http://dx.doi.org/10.1016/j.infsof.2007.09.004.

Beranek, G., Zuser, W., Grechenig, T., 2005. Functional group roles in software engineering teams. In: Proceedings of the 2005 Workshop on Human and Social Factors of Software Engineering. pp. 1–7. http://dx.doi.org/10.1145/1083106.1083108.

Boehm, B., 2006. A view of 20th and 21st century software engineering. In: Proceedings of the 28th International Conference on Software Engineering. pp. 12–29. http://dx.doi.org/10.1145/1134285.1134288.

Boehm, B., Turner, R., 2003a. Balancing Agility and Discipline: A Guide for the Perplexed. Addison-Wesley Longman Publishing Co. Inc., USA.

Boehm, B., Turner, R., 2003b. Using risk to balance agile and plan-driven methods. Computer 36 (6), 57–66. http://dx.doi.org/10.1109/MC.2003.1204376.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101. http://dx.doi.org/10.1191/1478088706QP063OA.

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M., Khalil, M., 2007. Lessons from applying the systematic literature review process within the software engineering domain. J. Syst. Softw. 80 (4), 571–583. http://dx.doi.org/10.1016/j.jss.2006.07.009.

Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica 1287–1294. http://dx.doi.org/10.2307/1911963.

Capiluppi, A., Fernandez-Ramil, J., Higman, J., Sharp, H.C., Smith, N., 2007. An empirical study of the evolution of an agile-developed software system. In: 29th International Conference on Software Engineering. ICSE'07, IEEE, pp. 511–518. http://dx.doi.org/10.1109/ICSE.2007.14.

Capretz, L.F., Ahmed, F., 2010. Why do we need personality diversity in software engineering? ACM SIGSOFT Softw. Eng. Not. 35 (2), 1–11. http://dx.doi.org/10.1145/1734103.1734111.

Cataldo, M., Ehrlich, K., 2012. The impact of communication structure on new product development outcomes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 3081–3090. http://dx.doi.org/10.1145/2207676.2208722.

Cataldo, M., Wagstrom, P.A., Herbsleb, J.D., Carley, K.M., 2006. Identification of coordination requirements: Implications for the design of collaboration and awareness tools. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. pp. 353–362. http://dx.doi.org/10.1145/1180875.1180929.

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: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society. ICSE-SEIS, IEEE, pp. 11–20. http://dx.doi.org/10.1109/ICSE-SEIS.2019.00010.

Centre for Reviews and Dissemination, 2016. Database of Attributes of Effects. University of York, Retrieved from http://www.york.ac.uk/crd/about/.

Chagas, A., Santos, M., Santana, C., Vasconcelos, A., 2015. The impact of human factors on agile projects. In: 2015 Agile Conference. IEEE, pp. 87–91. http://dx.doi.org/10.1109/Agile.2015.11.

Chang, K.T., Ehrlich, K., 2007. Out of sight but not out of mind? Informal networks, communication and media use in global software teams. In: Proceedings of the 2007 Conference of the Center for Advanced Studies on Collaborative Research. pp. 86–97. http://dx.doi.org/10.1145/1321211.1321221.

Chin, G., 2004. Agile Project Management: How to Succeed in the Face of Changing Project Requirements. AMACOM/American Management Association, Retrieved from http://www.mim.ac.mw/books/Chin's20Agile%20Project%20Management.pdf.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46. http://dx.doi.org/10.1177/001316446002000104.

Cohen, J., 1988. Statistical Power Analysis for the Social Sciences. Lawrence Erlbaum Associates, Hillsdale, NJ, Retrieved from https://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf.

Coram, M., Bohner, S., 2005. The impact of agile methods on software project management. In: 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems. ECBS'05, IEEE, pp. 363–370. http://dx.doi.org/10.1109/ECBS.2005.68.

Cruz, S., da Silva, F.Q., Capretz, L.F., 2015. Forty years of research on personality in software engineering: A mapping study. Comput. Hum. Behav. 46, 94–113. http://dx.doi.org/10.1016/j.chb.2014.12.008.

Cruzes, D.S., Dybå, T., 2011. Research synthesis in software engineering: A tertiary study. Inf. Softw. Technol. 53 (5), 440–455. http://dx.doi.org/10.1016/j.infsof.2011.01.004.

Cunha, A.B., Canen, A.G., Capretz, M.A., 2009. Personalities, cultures and software modeling: questions, scenarios and research directions. In: 2009 ICSE Workshop on Cooperative and Human Aspects on Software Engineering. IEEE, pp. 23–31. http://dx.doi.org/10.1109/CHASE.2009.5071406.

Damian, D., Izquierdo, L., Singer, J., Kwan, I., 2007. Awareness in the wild: Why communication breakdowns occur. In: International Conference on Global Software Engineering. ICGSE 2007, IEEE, pp. 81–90. http://dx.doi.org/10.1109/ICGSE.2007.13.

Damian, D., Kwan, I., Marczak, S., 2010. Requirements-driven collaboration: Leveraging the invisible relationships between requirements and people. In: Collaborative Software Engineering. Springer, Berlin, Heidelberg, pp. 57–76. http://dx.doi.org/10.1007/978-3-642-10294-3_3.

De Vries, R.E., Van den Hooff, B., De Ridder, J.A., 2006. Explaining knowledge sharing: The role of team communication styles, job satisfaction, and performance beliefs. Commun. Res. 33 (2), 115–135. http://dx.doi.org/10.1177/0093650205285366.

Denning, P.J., 2017. The forgotten engineer. Commun. ACM 60 (12), 20–23. http://dx.doi.org/10.1145/3152912.

Diebold, F.X., 1998. Elements of Forecasting. South-Western College Publishing, Cincinnati, OH, USA.

Dybå, T., Arisholm, E., Sjoberg, D.I., Hannay, J.E., Shull, F., 2007. Are two heads better than one? On the effectiveness of pair programming. IEEE Software 24 (6), 12–15. http://dx.doi.org/10.1109/MS.2007.158.

Dybå, T., Dingsøyr, T., 2008. Empirical studies of agile software development: A systematic review. Inf. Softw. Technol. 50 (9–10), 833–859. http://dx.doi.org/10.1016/j.infsof.2008.01.006.

Dyba, T., Dingsoyr, T., 2009. What do we know about agile software development? IEEE Software 26 (5), 6–9. http://dx.doi.org/10.1109/MS.2009.145.

Dyba, T., Dingsoyr, T., Hanssen, G.K., 2007. Applying systematic reviews to diverse study types: An experience report. In: First International Symposium on Empirical Software Engineering and Measurement. ESEM 2007, IEEE, pp. 225–234. http://dx.doi.org/10.1109/ESEM.2007.59.

Dybå, T., Kampenes, V.B., Sjøberg, D.I., 2006. A systematic review of statistical power in software engineering experiments. Inf. Softw. Technol. 48 (8), 745–755. http://dx.doi.org/10.1016/j.infsof.2005.08.009.

Ehrlich, K., Cataldo, M., 2012. All-for-one and one-for-all? A multi-level analysis of communication patterns and individual performance in geographically distributed software development. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. pp. 945–954. http://dx.doi.org/10.1145/2145204.2145345.

Fagerholm, F., Ikonen, M., Kettunen, P., Münch, J., Roto, V., Abrahamsson, P., 2014. How do software developers experience team performance in lean and agile environments? In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. pp. 1–10. http://dx.doi.org/10.1145/2601248.2601285.

Fatima, N., Nazir, S., Chuprat, S., 2019. Knowledge sharing, a key sustainable practice is on risk: An insight from modern code review. In: 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences. ICETAS, IEEE, pp. 1–6. http://dx.doi.org/10.1109/ICETAS48360.2019.9117444.

Feldt, R., Angelis, L., Torkar, R., Samuelsson, M., 2010. Links between the personalities, views and attitudes of software engineers. Inf. Softw. Technol. 52 (6), 611–624. http://dx.doi.org/10.1016/j.infsof.2010.01.001.

Fernández-Sanz, L., Misra, S., 2011. Influence of human factors in software quality and productivity. In: International Conference on Computational Science and Its Applications. Springer, Berlin, Heidelberg, pp. 257–269. http://dx.doi.org/10.1007/978-3-642-21934-4_22.

França, A.C.C., Gouveia, T.B., Santos, P.C., Santana, C.A., da Silva, F.Q., 2011. Motivation in software engineering: A systematic review update. In: 15th Annual Conference on Evaluation & Assessment in Software Engineering. EASE 2011, IET, pp. 154–163. http://dx.doi.org/10.1049/ic.2011.0019.

Gallivan, M.J., 2001. Striking a balance between trust and control in a virtual organization: A content analysis of open source software case studies. Inf. Syst. J. 11 (4), 277–304. http://dx.doi.org/10.1046/j.1365-2575.2001.00108.x.

Glantz, S.A., Slinker, B.K., 2001. Primer of Applied Regression & Analysis of Variance. McGraw-Hill, New York.

Glasziou, P.P., Irwig, L., Bain, C.J., Colditz, G.A., 2000. How to Review the Evidence: Systematic Identification and Review of the Scientific Literature. National Health & Medical Research Council, Canberra, Retrieved from https://www.nhmrc.gov.au/file/5741/download?token=4hv8FC-M.

Gomes, V., Marczak, S., 2012. Problems? We all know we have them. Do we have solutions too? A literature review on problems and their solutions in global software development. In: 2012 IEEE Seventh International Conference on Global Software Engineering. IEEE, pp. 154–158. http://dx.doi.org/10.1109/ICGSE.2012.43.

Gregory, P., Barroca, L., Taylor, K., Salah, D., Sharp, H., 2015. Agile challenges in practice: A thematic analysis. In: In International Conference on Agile Software Development. Springer, Cham, pp. 64–80. http://dx.doi.org/10.1007/978-3-319-18612-2_6.

Gujarati, D.N., 2003. Basic Econometrics Fourth Edition. Mcgraw-Hill, New York.

Guveyi, E., Aktas, M.S., Kalipsiz, O., 2020. Human factor on software quality: A systematic literature review. In: International Conference on Computational Science and Its Applications. Springer, Cham, pp. 918–930. http://dx.doi.org/10.1007/978-3-030-58811-3_65.

Hall, T., Baddoo, N., Beecham, S., Robinson, H., Sharp, H., 2009. A systematic review of theory use in studies investigating the motivations of software engineers. ACM Trans. Software Eng. Methodol. (TOSEM) 18 (3), 1–29. http://dx.doi.org/10.1145/1525880.1525883.

Hannay, J.E., Sjoberg, D.I., Dyba, T., 2007. A systematic review of theory use in software engineering experiments. IEEE Trans. Softw. Eng. 33 (2), 87–107. http://dx.doi.org/10.1109/TSE.2007.12.

Hanssen, G.K., Šmite, D., Moe, N.B., 2011. Signs of agile trends in global software engineering research: A tertiary study. In: 2011 IEEE Sixth International Conference on Global Software Engineering Workshop. IEEE, pp. 17–23. http://dx.doi.org/10.1109/ICGSE-W.2011.12.

Hawkins, D.M., 2004. The problem of overfitting. J. Chem. Inf. Comput. Sci. 44 (1), 1–12. http://dx.doi.org/10.1021/ci0342472.

Hayes, J.H., 2003. Do you like Pina Coladas? How improved communication can improve software quality. IEEE Software 20 (1), 90. http://dx.doi.org/10.1109/MS.2003.1159036.

Hazzan, O., Tomayko, J., 2004. Human aspects of software engineering: The case of extreme programming. In: International Conference on Extreme Programming and Agile Processes in Software Engineering. Springer, Berlin, Heidelberg, pp. 303–311. http://dx.doi.org/10.1007/978-3-540-24853-8_45.

Herbsleb, J.D., Mockus, A., Finholt, T.A., Grinter, R.E., 2001. An empirical study of global software development: Distance and speed. In: Proceedings of the 23rd International Conference on Software Engineering. ICSE 2001, IEEE, pp. 81–90. http://dx.doi.org/10.1109/ICSE.2001.919083.

Hidellaarachchi, D., Grundy, J., Hoda, R., Madampe, K., 2021. The effects of human aspects on the requirements engineering process: A systematic literature review. IEEE Trans. Softw. Eng. http://dx.doi.org/10.1109/TSE.2021.3051898.

Hoda, R., Salleh, N., Grundy, J., Tee, H.M., 2017. Systematic literature reviews in agile software development: A tertiary study. Inf. Softw. Technol. 85, 60–70. http://dx.doi.org/10.1016/j.infsof.2017.01.007.

Howison, J., Inoue, K., Crowston, K., 2006. Social dynamics of free and open source team communications. In: IFIP International Conference on Open Source Systems. Springer, Boston, MA, pp. 319–330. http://dx.doi.org/10.1007/0-387-34226-5_32.

Jia, J., Zhang, P., Capretz, L.F., 2016. Environmental factors influencing individual decision-making behavior in software projects: A systematic literature review. In: Proceedings of the 9th International Workshop on Cooperative and Human Aspects of Software Engineering. pp. 86–92. http://dx.doi.org/10.1145/2897586.2897589.

Jorgensen, M., Shepperd, M., 2006. A systematic review of software development cost estimation studies. IEEE Trans. Softw. Eng. 33 (1), 33–53. http://dx.doi.org/10.1109/TSE.2007.256943.

Kampenes, V.B., Dybå, T., Hannay, J.E., Sjøberg, D.I., 2007. A systematic review of effect size in software engineering experiments. Inf. Softw. Technol. 49 (11–12), 1073–1086. http://dx.doi.org/10.1016/j.infsof.2007.02.015.

Karn, J., Cowling, T., 2006. A follow up study of the effect of personality on the performance of software engineering teams. In: Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering. pp. 232–241. http://dx.doi.org/10.1145/1159733.1159769.

Kitchenham, B., 2004. Procedures for Performing Systematic Reviews. Keele University Technical Report TR/SE-0401,, Keele University, Keele, UK, July. Retrieved from https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf.

Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S., 2009. Systematic literature reviews in software engineering–A systematic literature review. Inf. Softw. Technol. 51 (1), 7–15. http://dx.doi.org/10.1016/j.infsof.2008.09.009.

Kitchenham, B., Budgen, D., Brereton, P., Turner, M., Charters, S., Linkman, S., 2007. Large-scale software engineering questions–expert opinion or empirical evidence? IET Software 1 (5), 161–171. http://dx.doi.org/10.1049/iet-sen:20060052.

Kitchenham, B., Charters, S., 2007a. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report., Retrieved from https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf.

Kitchenham, B., Charters, S., 2007b. Procedures for performing systematic literature reviews in software engineering. Keele University & Durham University, UK 33, 28.

Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O.P., Turner, M., Niazi, M., Linkman, S., 2010. Systematic literature reviews in software engineering–A tertiary study. Inf. Softw. Technol. 52 (8), 792–805. http://dx.doi.org/10.1016/j.infsof.2010.03.006.

Lee, D., Smith, A., Mortimer, M., 2011. Cultural differences affecting quality and productivity in Western/Asian offshore software development. In: Proceedings of the 3rd International Conference on Human Computer Interaction. pp. 29–39. http://dx.doi.org/10.1145/2407796.2407801.

Licorish, S.A., MacDonell, S.G., 2012. What affects team behavior? Preliminary linguistic analysis of communications in the Jazz repository. In: 2012 5th International Workshop on Co-Operative and Human Aspects of Software Engineering. CHASE, IEEE, pp. 83–89. http://dx.doi.org/10.1109/CHASE.2012.6223029.

Licorish, S., Philpott, A., MacDonell, S.G., 2009a. A prototype tool to support extended team collaboration in agile project feature management. In: Proceedings of the International Conference on Software Engineering Theory and Practice. SETP-09, ISRST, Orlando FL, USA, pp. 105–112, Retreived from http://orapp.aut.ac.nz/handle/10292/1742.

Licorish, S., Philpott, A., MacDonell, S.G., 2009b. Supporting agile team composition: A prototype tool for identifying personality (in) compatibilities. In: 2009 ICSE Workshop on Cooperative and Human Aspects on Software Engineering. IEEE, pp. 66–73. http://dx.doi.org/10.1109/CHASE.2009.5071413.

McHugh, M.L., 2012. Interrater reliability: The kappa statistic. Biochemia Med. 22 (3), 276–282, Retrieved from https://hrcak.srce.hr/89395.

Meldrum, S., Licorish, S.A., Savarimuthu, B.T.R., 2017. Crowdsourced knowledge on stack overflow: A systematic mapping study. In: Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering. pp. 180–185. http://dx.doi.org/10.1145/3084226.3084267.

Meyer, A.N., 2018. Fostering software developers' productivity at work through self-monitoring and goal-setting. In: 2018 IEEE/ACM 40th International Conference on Software Engineering: Companion. ICSE-Companion, IEEE, pp. 480–483. http://dx.doi.org/10.1145/3183440.3183446.

Mohanani, R., Salman, I., Turhan, B., Rodríguez, P., Ralph, P., 2018. Cognitive biases in software engineering: A systematic mapping study. IEEE Trans. Softw. Eng. 46 (12), 1318–1339. http://dx.doi.org/10.1109/TSE.2018.2877759.

Mujtaba, S., Petersen, K., Feldt, R., Mattsson, M., 2008. Software Product Line Variability: A Systematic Mapping Study. School of Engineering, Blekinge Inst. of Technology, Retrieved from https://www.researchgate.net/profile/Michael-Mattsson/publication/228570142_Software_Product_Line_Variability_A_Systematic_Mapping_Study/links/54d0a8e80cf20323c218713c/Software-Product-Line-Variability-A-Systematic-Mapping-Study.pdf.

Ocker, R.J., Fjermestad, J., 2008. Communication differences in virtual design teams: Findings from a multi-method analysis of high and low performing experimental teams. ACM SIGMIS Database: Database Adv. Inf. Syst. 39 (1), 51–67. http://dx.doi.org/10.1145/1341971.1341977.

Osborne, J.W., Waters, E., 2002. Four assumptions of multiple regression that researchers should always test. Pract. Assess., Res. Eval. 8 (1), 2. http://dx.doi.org/10.1080/00401706.1991.10484839.

Paul, S.R., Fung, K.Y., 1991. A generalized extreme studentized residual multiple-outlier-detection procedure in linear regression. Technometrics 33 (3), 339–348.

Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M., 2008. Systematic mapping studies in software engineering. In: 12th International Conference on Evaluation and Assessment in Software Engineering. EASE, pp. 1–10. http://dx.doi.org/10.14236/ewic/EASE2008.8.

Petersen, K., Vakkalanka, S., Kuzniarz, L., 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. Inf. Softw. Technol. 64, 1–18. http://dx.doi.org/10.1016/j.infsof.2015.03.007.

Rajendran, M., 2005. Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin self-perception inventory. J. Manag. Dev. 24, 738–753. http://dx.doi.org/10.1108/02621710510613753.

Ropponen, J., Lyytinen, K., 2000. Components of software development risk: How to address them? A project manager survey. IEEE Trans. Softw. Eng. 26 (2), 98–112. http://dx.doi.org/10.1109/32.841112.

Sach, R., Sharp, H., Petre, M., 2011. Software engineers' perceptions of factors in motivation: The work, people, obstacles. In: 2011 International Symposium on Empirical Software Engineering and Measurement. IEEE, pp. 368–371. http://dx.doi.org/10.1109/ESEM.2011.50.

Scacchi, W., 2002. Process models in software engineering. In: Encyclopedia of Software Engineering. http://dx.doi.org/10.1002/0471028959.sof250.

Schmidt, R., Lyytinen, K., Keil, M., Cule, P., 2001. Identifying software project risks: An international Delphi study. J. Manage. Inf. Syst. 17 (4), 5–36. http://dx.doi.org/10.1080/07421222.2001.11045662.

Schober, P., Boer, C., Schwarte, L.A., 2018. Correlation coefficients: Appropriate use and interpretation. Anesth. Analg. 126 (5), 1763–1768. http://dx.doi.org/10.1213/ANE.0000000000002864.

Sfetsos, P., Stamelos, I., Angelis, L., Deligiannis, I., 2006. Investigating the impact of personality types on communication and collaboration-viability in pair programming–An empirical study. In: International Conference on Extreme Programming and Agile Processes in Software Engineering. Springer, Berlin, Heidelberg, pp. 43–52. http://dx.doi.org/10.1007/11774129_5.

Sharma, V.S., Kaulgud, V., 2011. Studying team evolution during software testing. In: Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering. pp. 72–75. http://dx.doi.org/10.1145/1984642.1984660.

Shaw, M., 2009. Continuing prospects for an engineering discipline of software. IEEE Software 26 (6), 64–67. http://dx.doi.org/10.1109/MS.2009.172.

Siddiqui, M.S., Hussain, S.J., 2006. Comprehensive software development model. In: IEEE International Conference on Computer Systems and Applications, 2006. IEEE, pp. 353–360. http://dx.doi.org/10.1109/AICCSA.2006.205113.

Sjoberg, D.I., Dyba, T., Jorgensen, M., 2007. The future of empirical methods in software engineering research. In: Future of Software Engineering. FOSE'07, IEEE, pp. 358–378. http://dx.doi.org/10.1109/FOSE.2007.30.

Slinker, B.K., Glantz, S.A., 2008. Multiple linear regression: Accounting for multiple simultaneous determinants of a continuous dependent variable. Circulation 117 (13), 1732–1737. http://dx.doi.org/10.1161/CIRCULATIONAHA.106.654376.

Stavru, S., 2014. A critical examination of recent industrial surveys on agile method usage. J. Syst. Softw. 94, 87–97. http://dx.doi.org/10.1016/j.jss.2014.03.041.

Svahnberg, M., Gorschek, T., Feldt, R., Torkar, R., Saleem, S.B., Shafique, M.U., 2010. A systematic review on strategic release planning models. Inf. Softw. Technol. 52 (3), 237–248. http://dx.doi.org/10.1016/j.infsof.2009.11.006.

The Standish Group, 1994. The chaos report. Retrieved from http://www.standishgroup.com/sample_research/PDFpages/extreme_chaos.pdf.

The Standish Group, 2001. Extreme chaos. Retrieved from http://www.standishgroup.com/sample_research/PDFpages/extreme_chaos.pdf.

Thompson, C.G., Kim, R.S., Aloe, A.M., Becker, B.J., 2017. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic Appl. Soc. Psychol. 39 (2), 81–90. http://dx.doi.org/10.1080/01973533.2016.1277529.

Unterkalmsteiner, M., Gorschek, T., Islam, A.M., Cheng, C.K., Permadi, R.B., Feldt, R., 2011. Evaluation and measurement of software process improvement—A systematic literature review. IEEE Trans. Softw. Eng. 38 (2), 398–424. http://dx.doi.org/10.1109/TSE.2011.26.

Varona, D., Capretz, L.F., Piñero, Y., Raza, A., 2012. Evolution of software engineers' personality profile. ACM SIGSOFT Software Eng. Not. 37 (1), 1–5. http://dx.doi.org/10.1145/2088883.2088901.

Walz, D.B., Elam, J.J., Curtis, B., 1993. Inside a software design team: Knowledge acquisition, sharing, and integration. Commun. ACM 36 (10), 63–77. http://dx.doi.org/10.1145/163430.163447.

Webster, J., Watson, R.T., 2002. Analyzing the past to prepare for the future: Writing a literature review. MIS Q. xiii–xxiii, Retrieved from http://www.jstor.org/stable/4132319.

Wirth, N., 2008. A brief history of software engineering. IEEE Ann. Hist. Comput. 30 (3), 32–39. http://dx.doi.org/10.1109/MAHC.2008.33.

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. pp. 1–10. http://dx.doi.org/10.1145/2601248.2601268.

Wynekoop, J.L., Walz, D.B., 2000. Investigating traits of top performing software developers. Inf. Technol. People 13 (3), 186–195. http://dx.doi.org/10.1108/09593840010377626.

Yaseen, M., Baseer, S., Sherin, S., 2015. Critical challenges for requirement implementation in context of global software development: A systematic literature review. In: 2015 International Conference on Open Source Systems & Technologies. ICOSST, IEEE, pp. 120–125. http://dx.doi.org/10.1109/ICOSST.2015.7396413.

Zhang, H., Babar, M.A., Tell, P., 2011. Identifying relevant studies in software engineering. Inf. Softw. Technol. 53 (6), 625–637. http://dx.doi.org/10.1016/j.infsof.2010.12.010.

Zheng, J., Veinott, E., Bos, N., Olson, J.S., Olson, G.M., 2002. Trust without touch: Jumpstarting long-distance trust with initial social activities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 141–146. http://dx.doi.org/10.1145/503376.503402.

Zhou, M., Mockus, A., 2011. Does the initial environment impact the future of developers? In: Proceedings of the 33rd International Conference on Software Engineering. pp. 271–280. http://dx.doi.org/10.1145/1985793.1985831.

Zolduoarrati, E., Licorish, S., Stanger, N., 2022. Human aspects tertiary dataset. http://dx.doi.org/10.5281/zenodo.7444414.

Zwikael, O., Ahn, M., 2011. The effectiveness of risk management: An analysis of project risk planning across industries and countries. Risk Anal.: Int. J. 31 (1), 25–37. http://dx.doi.org/10.1111/j.1539-6924.2010.01470.x.

Reviews included in the tertiary study

Abrar, M.F., Khan, M.S., Ali, S., Ali, U., Majeed, M.F., Ali, A., et al., 2019. Motivators for large-scale agile adoption from management perspective: A systematic literature review. IEEE Access 7, 22660–22674. http://dx.doi.org/10.1109/ACCESS.2019.2896212.

Ali, S., Hongqi, L., Khan, S.U., Zhongguo, Y., Liping, Z., 2017. Success factors for software outsourcing partnership management: An exploratory study using systematic literature review. IEEE Access 5, 23589–23612. http://dx.doi.org/10.1109/ACCESS.2017.2764946.

Ali, S., Khan, S.U., 2014. Critical success factors for software outsourcing partnership (SOP): A systematic literature review. In: 2014 IEEE 9th International Conference on Global Software Engineering. IEEE, pp. 153–162. http://dx.doi.org/10.1109/ICGSE.2014.12.

Alzoubi, Y.I., Gill, A.Q., Al-Ani, A., 2016. Empirical studies of geographically distributed agile development communication challenges: A systematic review. Inf. Manag. 53 (1), 22–37. http://dx.doi.org/10.1016/j.im.2015.08.003.

Anchundia, C.E., Fonseca, E.R., 2020. Resources for reproducibility of experiments in empirical software engineering: Topics derived from a secondary study. IEEE Access 8, 8992–9004. http://dx.doi.org/10.1109/ACCESS.2020.2964587.

Anh, N.D., Cruzes, D.S., Conradi, R., 2012. Dispersion, coordination and performance in global software teams: A systematic review. In: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. pp. 129–138. http://dx.doi.org/10.1145/2372251.2372274.

Anu, V., Hu, W., Carver, J.C., Walia, G.S., Bradshaw, G., 2018. Development of a human error taxonomy for software requirements: A systematic literature review. Inf. Softw. Technol. 103, 112–124. http://dx.doi.org/10.1016/j.infsof.2018.06.011.

Anwar, R., Rehman, M., Wang, K.S., Hashmani, M.A., 2019. Systematic literature review of knowledge sharing barriers and facilitators in global software development organizations using concept maps. IEEE Access 7, 24231–24247. http://dx.doi.org/10.1109/ACCESS.2019.2895690.

Barroso, A.S., Madureira, J.S., Soares, M.S., do Nascimento, R.P., 2017. Influence of human personality in software engineering-A systematic literature review. In: International Conference on Enterprise Information Systems, vol. 2, SciTePress, pp. 53–62. http://dx.doi.org/10.5220/0006292000530062.

Brito, dos Santos K., de Lima, A.A., Ferreira, S.E., de Arruda Burégio, V., Garcia, V.C., de Lemos Meira, S.R., 2020. Evolution of the web of social machines: A systematic review and research challenges. IEEE Trans. Comput. Soc. Syst. 7 (2), 373–388. http://dx.doi.org/10.1109/TCSS.2019.2961269.

Britto, R., Freitas, V., Mendes, E., Usman, M., 2014. Effort estimation in global software development: A systematic literature review. In: 2014 IEEE 9th International Conference on Global Software Engineering. IEEE, pp. 135–144. http://dx.doi.org/10.1109/ICGSE.2014.11.

Canedo, E.D., Santos, G.A., 2019. Factors affecting software development productivity: An empirical study. In: Proceedings of the XXXIII Brazilian Symposium on Software Engineering. pp. 307–316. http://dx.doi.org/10.1145/3350768.3352491.

Chagas, L.F., Carvalho, D.D.D., Lima, A.M., Reis, C.A.L., 2014. Systematic literature review on the characteristics of agile project management in the context of maturity models. In: International Conference on Software Process Improvement and Capability Determination. Springer, Cham, pp. 177–189. http://dx.doi.org/10.1007/978-3-319-13036-1_16.

Ciupe, A., Meza, S., Ionescu, R., Orza, B., 2017. Practical agile in higher education: A systematic mapping study. In: 2017 XXVI International Conference on Information, Communication and Automation Technologies. ICAT, IEEE, pp. 1–6. http://dx.doi.org/10.1109/ICAT.2017.8171626.

Costa, V.G., França, C., 2019. How workspaces influence software development? Preliminary results of a systematic literature review. In: Anais Do VII Workshop de Visualização, Evolução E Manutenção de Software. SBC, pp. 41–48. http://dx.doi.org/10.5753/vem.2019.7584.

Cruz, S.S., da Silva, F.Q., Monteiro, C.V., Santos, P., Rossilei, I., dos Santos, M.T., 2011. Personality in software engineering: Preliminary findings from a systematic literature review. In: 15th Annual Conference on Evaluation & Assessment in Software Engineering. EASE 2011, IET, pp. 1–10. http://dx.doi.org/10.1049/ic.2011.0001.

da Silva, F.Q., Costa, C., Franca, A.C.C., Prikladinicki, R., 2010. Challenges and solutions in distributed software development project management: A systematic literature review. In: 2010 5th IEEE International Conference on Global Software Engineering. IEEE, pp. 87–96. http://dx.doi.org/10.1109/ICGSE.2010.18.

de Barros Sampaio, S.C., Barros, E.A., de Aquino, G.S., e Silva, M.J.C., de Lemos Meira, S.R., 2010. A review of productivity factors and strategies on software development. In: 2010 Fifth International Conference on Software Engineering Advances. IEEE, pp. 196–204. http://dx.doi.org/10.1109/ICSEA.2010.37.

Durán, M., Juárez-Ramírez, R., Jiménez, S., Tona, C., 2019. Taxonomy for complexity estimation in agile methodologies: A systematic literature review. In: 2019 7th International Conference in Software Engineering Research and Innovation. CONISOFT, IEEE, pp. 87–96. http://dx.doi.org/10.1109/CONISOFT.2019.00022.

Dutra, A.C., Prikladnicki, R., França, C., 2015. What do we know about high performance teams in software engineering? Results from a systematic literature review. In: 2015 41st Euromicro Conference on Software Engineering and Advanced Applications. IEEE, pp. 183–190. http://dx.doi.org/10.1109/SEAA.2015.24.

El Bajta, M., Idri, A., 2020. Identifying software cost attributes of software project management in global software development: An integrative framework. In: Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications. pp. 1–5. http://dx.doi.org/10.1145/3419604.3419780.

Fitriani, W.R., Rahayu, P., Sensuse, D.I., 2016. Challenges in agile software development: A systematic literature review. In: 2016 International Conference on Advanced Computer Science and Information Systems. ICACSIS, IEEE, pp. 155–164. http://dx.doi.org/10.1109/ICACSIS.2016.7872736.

Flores, A.P.M., de Alencar, F.M.R., 2020. Competencies development based on thinking-based learning in software engineering: An action-research. In: Proceedings of the 34th Brazilian Symposium on Software Engineering. pp. 680–689. http://dx.doi.org/10.1145/3422392.3422481.

Garousi, V., Petersen, K., Ozkan, B., 2016. Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review. Inf. Softw. Technol. 79, 106–127. http://dx.doi.org/10.1016/j.infsof.2016.07.006.

Ghanbari, H., Vartiainen, T., Siponen, M., 2018. Omission of quality software development practices: A systematic literature review. ACM Comput. Surv. 51 (2), 1–27. http://dx.doi.org/10.1145/3177746.

Groeneveld, W., Vennekens, J., Aerts, K., 2019. Software engineering education beyond the technical: A systematic literature review. http://dx.doi.org/10.48550/arXiv.1910.09865, arXiv preprint arXiv:1910.09865.

Heaton, D., Carver, J.C., 2015. Claims about the use of software engineering practices in science: A systematic literature review. Inf. Softw. Technol. 67, 207–219. http://dx.doi.org/10.1016/j.infsof.2015.07.011.

Hummel, M., Rosenkranz, C., Holten, R., 2013. The role of communication in agile systems development. Bus. Inf. Syst. Eng. 5 (5), 343–355. http://dx.doi.org/10.1007/s12599-013-0282-4.

Jabbari, R., bin Ali, N., Petersen, K., Tanveer, B., 2016. What is DevOps? A systematic mapping study on definitions and practices. In: Proceedings of the Scientific Workshop Proceedings of XP2016. pp. 1–11. http://dx.doi.org/10.1145/2962695.2962707.

Khan, A.A., Keung, J., Hussain, S., Niazi, M., Tamimy, M.M.I., 2017. Understanding software process improvement in global software development: A theoretical framework of human factors. ACM SIGAPP Appl. Comput. Rev. 17 (2), 5–15. http://dx.doi.org/10.1145/3131080.3131081.

Komal, B., Janjua, U.I., Madni, T.M., 2019. Identification of scope creep factors and their impact on software project success. In: In 2019 International Conference on Computer and Information Sciences. ICCIS, IEEE, pp. 1–5. http://dx.doi.org/10.1109/ICCISci.2019.8716390.

Kroll, J., Hashmi, S.I., Richardson, I., Audy, J.L., 2013. A systematic literature review of best practices and challenges in follow-the-sun software development. In: 2013 IEEE 8th International Conference on Global Software Engineering Workshops. IEEE, pp. 18–23. http://dx.doi.org/10.1109/ICGSEW.2013.10.

Lenberg, P., Feldt, R., Wallgren, L.G., 2014. Towards a behavioral software engineering. In: Proceedings of the 7th International Workshop on Cooperative and Human Aspects of Software Engineering. pp. 48–55. http://dx.doi.org/10.1145/2593702.2593711.

Lenberg, P., Feldt, R., Wallgren, L.G., 2015. Behavioral software engineering: A definition and systematic literature review. J. Syst. Softw. 107, 15–37. http://dx.doi.org/10.1016/j.jss.2015.04.084.

Lima, A.M., Reis, R.Q., de Souza, C.R., 2019. Using a systematic literature review to strengthen the evidence supporting a simulation model of distributed software projects. In: 2019 45th Euromicro Conference on Software Engineering and Advanced Applications. SEAA, IEEE, pp. 371–378. http://dx.doi.org/10.1109/SEAA.2019.00063.

Matturro, G., Raschetti, F., Fontán, C., 2019. A systematic mapping study on soft skills in software engineering. J. UCS 25 (1), 16–41. http://dx.doi.org/10.3217/jucs-025-01-0016.

Melo, C.D.O., Santana, C., Kon, F., 2012. Developers motivation in agile teams. In: 2012 38th Euromicro Conference on Software Engineering and Advanced Applications. IEEE, pp. 376–383. http://dx.doi.org/10.1109/SEAA.2012.45.

Nazir, S., Fatima, N., Chuprat, S., 2019. Situational factors affecting software engineers sustainability: A vision of modern code review. In: 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences. ICETAS, IEEE, pp. 1–6. http://dx.doi.org/10.1109/ICETAS48360.2019.9117366.

Onoue, S., Hata, H., Kula, R.G., Matsumoto, K., 2018. Human capital in software engineering: A systematic mapping of reconceptualized human aspect studies. http://dx.doi.org/10.48550/arXiv.1805.03844, arXiv preprint arXiv:1805.03844.

Pedreira, O., García, F., Brisaboa, N., Piattini, M., 2015. Gamification in software engineering–A systematic mapping. Inf. Softw. Technol. 57, 157–168. http://dx.doi.org/10.1016/j.infsof.2014.08.007.

Pirzadeh, L., 2010. Human Factors in Software Development: A Systematic Literature Review (Master's thesis). Chalmers University of Technology, doi:20.500.12380/126748.

Ramírez-Mora, S.L., Oktaba, H., 2017. Productivity in agile software development: A systematic mapping study. In: 2017 5th International Conference in Software Engineering Research and Innovation. CONISOFT, IEEE, pp. 44–53. http://dx.doi.org/10.1109/CONISOFT.2017.00013.

Sánchez-Gordón, M., Colomo-Palacios, R., 2019. Taking the emotional pulse of software engineering—A systematic literature review of empirical studies. Inf. Softw. Technol. 115, 23–43. http://dx.doi.org/10.1016/j.infsof.2019.08.002.

Schneider, S., Torkar, R., Gorschek, T., 2013. Solutions in global software engineering: A systematic literature review. Int. J. Inf. Manage. 33 (1), 119–132. http://dx.doi.org/10.1016/j.ijinfomgt.2012.06.002.

Senapathi, M., Srinivasan, A., 2013. Sustained agile usage: A systematic literature review. In: Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering. pp. 119–124. http://dx.doi.org/10.1145/2460999.2461016.

Shameem, M., Chandra, B., Kumar, R.R., Kumar, C., 2018. A systematic literature review to identify human related challenges in globally distributed agile software development: Towards a hypothetical model for scaling agile methodologies. In: 2018 4th International Conference on Computing Communication and Automation. ICCCA, IEEE, pp. 1–7. http://dx.doi.org/10.1109/CCAA.2018.8777533.

Shameem, M., Kumar, C., Chandra, B., 2017. Challenges of management in the operation of virtual software development teams: A systematic literature review. In: 2017 4th International Conference on Advanced Computing and Communication Systems. ICACCS, IEEE, pp. 1–8. http://dx.doi.org/10.1109/ICACCS.2017.8014695.

Sharp, H., Baddoo, N., Beecham, S., Hall, T., Robinson, H., 2009. Models of motivation in software engineering. Inf. Softw. Technol. 51 (1), 219–233. http://dx.doi.org/10.1016/j.infsof.2008.05.009.

Silveira, K.K., Prikladnicki, R., 2019. A systematic mapping study of diversity in software engineering: A perspective from the agile methodologies. In: 2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering. CHASE, IEEE, pp. 7–10. http://dx.doi.org/10.1109/CHASE.2019.00010.

Sinha, R., Shameem, M., Kumar, C., 2020. SWOT: Strength, weaknesses, opportunities, and threats for scaling agile methods in global software development. In: Proceedings of the 13th Innovations in Software Engineering Conference on Formerly Known As India Software Engineering Conference. pp. 1–10. http://dx.doi.org/10.1145/3385032.3385037.

Sirazitdinov, I., Succi, G., Ivanov, V., 2020. A systematic literature review of studies related to mental activities of software developers. In: 2020 International Conference Nonlinearity, Information and Robotics. NIR, IEEE, pp. 1–11. http://dx.doi.org/10.1109/NIR50484.2020.9290239.

Soomro, A.B., Salleh, N., 2014. A systematic review of the effects of team climate on software team productivity. In: Asia-Pacific World Congress on Computer Science and Engineering. IEEE, pp. 1–7. http://dx.doi.org/10.1109/APWCCSE.2014.7053876.

Suali, A.J., Fauzi, S.S.M., Sobri, W.A.W.M., Nasir, M.H.N.M., 2017. Developers' coordination issues and its impact on software quality: A systematic review. In: 2017 3rd International Conference on Science in Information Technology. ICSITech, IEEE, pp. 659–663. http://dx.doi.org/10.1109/ICSITech.2017.8257195.

Uludag, Ö., Kleehaus, M., Caprano, C., Matthes, F., 2018. Identifying and structuring challenges in large-scale agile development based on a structured literature review. In: 2018 IEEE 22nd International Enterprise Distributed Object Computing Conference. EDOC, IEEE, pp. 191–197. http://dx.doi.org/10.1109/EDOC.2018.00032.

Veras, N.L., Rocha, L.S., Viana, W., 2020. Flipped classroom in software engineering: A systematic mapping study. In: Proceedings of the 34th Brazilian Symposium on Software Engineering. pp. 720–729. http://dx.doi.org/10.1145/3422392.3422490.

Vishnubhotla, S.D., Mendes, E., Lundberg, L., 2018. An insight into the capabilities of professionals and teams in agile software development: A systematic literature review. In: Proceedings of the 2018 7th International Conference on Software and Computer Applications. pp. 10–19. http://dx.doi.org/10.1145/3185089.3185096.

Walia, G.S., Carver, J.C., 2009. A systematic literature review to identify and classify software requirement errors. Inf. Softw. Technol. 51 (7), 1087–1109. http://dx.doi.org/10.1016/j.infsof.2009.01.004.

Elijah Zolduoarrati is a Research Assistant and PhD student in the Department of Information Science at the University of Otago, New Zealand, where he also earned his BAppSc and MSc. His research portfolio covers applied data science, big data, bioinformatics, cloud computing, computational linguistics, cyber security, data (analysis, architecture, brokerage, cleansing, exploration, inspection, management, mining, modelling, predictions, preservation, reproducibility, transformation and visualisation), diversity (gender inclusivity and equality), education, human–computer interaction, human factors, information assurance, internet of things, machine learning, micromarketing, operating systems, quantitative and qualitative research methods and techniques, robotics, social and behavioural sciences also software engineering.

Sherlock A. Licorish is a Senior Lecturer in the Department of Information Science at University of Otago, New Zealand. He was awarded his PhD by Auckland University of Technology (AUT), and joined University of Otago in 2014. His research portfolio covers agile methodologies, practices and processes, teams and human factors, human computer interaction, research methods and techniques, software code quality, static analysis tools, machine learning applications and software analytics. Sherlock occupies several service roles across the university, nationally and internationally.

Nigel Stanger is a Lecturer in the Department of Information Science at the University of Otago, New Zealand, where he also earned his PhD. His research interests include applied data science (focusing particularly on the software engineering domain), data analysis reproducibility, data management and preservation, human factors in software engineering, defect prediction, and computing education. He has served on the committees of EASE, eScience, ICCE, and APCCM, and was the general co-chair of the 2021 Australasian Computer Science Week multi-conference. He is a member of ACM and IT Professionals New Zealand.