The Journal of Systems & Software 204 (2023) 111795

Contents lists available at ScienceDirect

The Journal of Systems & Software

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

A systematic mapping study on group work research in computing education projects☆

Mikko Kokkoniemi *, Ville Isomöttönen

Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

A R T I C L E I N F O

Article history:
Received 2 December 2022
Received in revised form 27 April 2023
Accepted 25 June 2023
Available online 28 June 2023

Keywords:
Systematic mapping study
Group work
Group projects
Project education
PjBL

A B S T R A C T

Context: For developing students' group- and teamwork skills needed in the team-oriented work environments of the software industry, the role of project-based learning is considered central. Yet there does not appear to be a proper mapping of the current group work research in the computer science project education literature. Thus, the current state of group work research in the research area is somewhat unknown.
Objective: This study aims to form an overview of how research has addressed students' group work in the field of computer science (CS) and software engineering (SE) to identify research gaps as well as suitable topics for more detailed literature reviews.
Methods: A systematic mapping study was used to investigate how group work in tertiary education has been undertaken in the literature during the past decade.
Results: Based on the selected papers, the most investigated group work areas were related to the assessment of groups, group formation, communication, and cooperation. The research appeared to be quite narrowly focused on a few areas. Most of the papers were experience or evaluation research. A case study using interviews or questionnaires to gather data from a single course was the most representative type of study. The papers were mainly published in scientific conferences. The use of theoretical frameworks was limited, with a focus on a few established frameworks. Tuckman's group development theory was the predominant framework, while other commonly used concepts and theories include social loafing, Kolb's learning style theory, and the Big Five personality traits model.
Conclusion: Out of 7515 papers screened, 225 were deemed eligible and analyzed. We conclude a need for more focused group work research in CS/SE student projects, in which education is inspected from particular perspectives. This would create identifiable lines of research and structure the research area. Relatedly, we suggest that the underused theoretical frameworks can inspire important research: group interventions would benefit from a socially shared regulated learning perspective, explicit use of justice theories would improve theoretical understandings of group behavior, and transactional distance would help analyze how students adopt a software process. Moreover, the research area could be precipitated by novel theoretical perspectives. For practitioners, those implementing a group project course can benefit from a large amount of literature on assessment and group formation, which are issues on which the teacher must take a position. We also include a lessons-learned summary for teachers. Generally, the present results outlining the field in a structured way can facilitate research-based teaching.

1. Introduction

Project courses play a key role in computer science (CS) and software engineering (SE) curricula. According to Tomayko (1998), projects have been a permanent part of computing education since the first documented curriculum in 1968. As a teaching method they offer a learning environment that models real-world situations and provides students with an opportunity to demonstrate professional skills (Majanoja and Vasankari, 2018; Fincher et al., 2001; Clear et al., 2001). Additionally, projects act as a vehicle for effective learning (Fincher et al., 2001). Clear et al. (2001) describe a project course as a "culminating and integrative educational experience", which "places unique demands on students". Despite the projects' demanding nature, students enjoy

☆ Editor: Prof W. Eric Wong.

* Correspondence to: University of Jyväskylä, Faculty of Information Technology, P.O. Box 35, FI 40014, Finland.
E-mail address: mikko@kokkoniemi.com (M. Kokkoniemi).

https://doi.org/10.1016/j.jss.2023.111795

the practical approach and see it benefiting their subsequent careers (Isomöttönen, 2011). The form and the arrangement of a project course can vary a lot, which can be observed in taxonomies published by Fincher et al. (2001) and Clear et al. (2001), and more recently by Tenhunen et al. (2023).

Although projects can be performed individually, one typical method to prepare students for working life is a group project (e.g., Jun, 2010; Pérez and Rubio, 2020) Still, the current literature regularly presents observations about the problems that employers face because of the graduates' lacking group- and/or collaboration skills as a justification for the study (e.g., Figl, 2010; Marques and Ochoa, 2014; Scott et al., 2019). There seems to be a gap between employers' expectations and graduates' level of expertise (Hernández-March et al., 2009), with some industry consultations showing that teamwork skills can even be seen as more crucial than development skills (Barr and Parkinson, 2019). In another study covering engineering fields, academics emphasized less teamwork skills than employers but yet part of the data showed acknowledgment of teamwork as an important future goal (Khoo et al., 2020). These observations imply a need to gain a more comprehensive understanding of the group work phenomena in CS and SE project education contexts than we currently have.

Provocatively, one may even ask: how clear the picture of group work phenomena in project education is currently in CS and SE literature? The validity of the question looks apparent when we consider the results of a systematic review of literature on global software engineering courses (Clear et al., 2015) that identified group work as an area requiring attention. In line with the previous observations in the field (Isomöttönen, 2011), our experience is that evaluating the current understanding of group work in CS/SE project education is not trivial because the amount of research on project-based learning has grown vast during the past years and it has become burdensome to perceive the overall situation.

We conducted a systematic mapping study investigating how group work in tertiary education has been addressed in the CS and SE project-based learning literature over the past decade. We used a research method described by Petersen et al. (2008, 2015), which is a useful method for objectively summarizing and classifying information concerning a research question. In line with the common goals of a systematic mapping study (Petersen et al., 2008), we aim to provide an overview of the relevant literature in the research area and identify research gaps and potential topics for further systematic literature reviews. This kind of overview of the current research interests in a particular field is a useful tool for both the practitioners and the students entering the field (Isomöttönen et al., 2018). We searched for related systematic literature studies and concluded a research gap for this larger-scale mapping focus. The related literature studies appeared to have a specific focus or limited educational focus, and are reviewed in Section 2.4.

Our research data consisted of conference and journal papers published between 2010–2021. We believe that this time frame is appropriate for illustrating the nature of current research, without generating an excessive amount of data that would be challenging to handle within a single research project. However, to provide context for the reader, we provide a historical overview of research on group work in project-based learning in Sections 2.1–2.3.

As is typical for systematic mapping studies (Petersen et al., 2015), our study included both topic-independent and topic-dependent research questions. Topic-independent research questions, such as determining the time, forum and type of the publication, enable comparing the results with other systematic mappings. Additionally, we collected information on the geographical locations of the authors, which, together with the forum, clearly describes where the papers in the research area have been conducted. The topic-dependent questions looked into research methods, theoretical frameworks, and group work areas of the inspected literature. Given that group work is an intricate phenomenon that can stimulate a wide variety of research interests, these questions shed light on what has been studied and if some topics and methods are more common than others.

2. Related work

2.1. The project method

The literature often identifies William H. Kilpatrick as the coiner of the project method—Kilpatrick introduced project method as a new teaching philosophy in his famous paper in 1918 (Kilpatrick, 1918). However, Knoll (2012) questioned Kilpatrick's contribution and traced the origins of the project method to Europe where architect students had demonstrated their skills with design activities for real projects two centuries earlier. Knoll (2012) explained that the project method extended to engineering fields in Europe and US soon after this. Additionally, it can be observed that the early project education activities described by Knoll appeared to have a goal frequently stated in SE project course reports during the past decades: exposing students to realistic experiences that are relevant for them after graduation (e.g., Knoke, 1991; Johansson and Molin, 1996; Daniels and Asplund, 1999; Christensen and Rundus, 2003; Bothe, 2001; Mochol and Tolksdorf, 2009).

One dated but still useful definition is available from Adderley. Adderley et al. (1975) explained that the project method includes a problem that is often but not necessarily identified by students, emphasizes student initiative with teachers in an advisory role, produces a concrete outcome such as a thesis, report, plan, or computer program, and requires a considerably long period of work through one or more phases. Later, Helle and Tynjälä (2006) accordingly emphasized the authentic problems and concrete outcomes as the key aspects of project-based learning, while noting that problem-based learning – a sibling of project-based learning – can be based on paper cases.

The Adderley's definition can be complemented by categorizing approaches to project-based learning. For instance, Morgan (1983) divided project education into three categories – project exercise, project component, and project orientation – based on how prevailing role the project-based approach takes. The project exercise requires students to apply previously learned content, whereas the project component indicates an assignment with larger goals and real-life authenticity, and tends to present a multidisciplinary challenge. The project orientation indicates a project-based philosophy being applied to the whole program, in which case the project-based approach underpins all education and direct instruction is offered only to complement project-based offerings.

2.2. Project education in CS and SE curricula

Several taxonomic works have identified characteristics and design attributes of CS and SE project courses (e.g., Shaw and Tomayko, 1991; Knoke, 1991; Fincher et al., 2001; Clear et al., 2001; Burge and Gannod, 2009; Tenhunen et al., 2023) and thereby define project-based learning in a disciplinary context. Shaw and Tomayko (1991, p. 7) considered project size and differentiated between a "small group project", "large project team", and "project only" depending on how substantial a role project work took compared to discussion and lecturing during a course, and Knoke (1991) complemented this scheme by proposing a medium-sized project. These characterizations appear comparable to the Morgan's project exercise and project component above. Fincher et al. (2001) identified for example projects with research characteristics, projects focusing on designing and building products, projects emphasizing process, projects involving industry and real clients, projects performed individually or in groups, and integrative capstone projects. Clear et al. (2001) discussed several design aspects in detail, including topics such as curricular goals, group formation, team size, types of deliverables, use of sponsors, and role of nondisclosure agreements. Burge and Gannod (2009) differentiated between degrees of realism in students' work, noting for instance that all groups working on the same project is a less realistic setting than each group producing a working system and that a realistic project affords experiences of requirements elicitation. Tenhunen et al. (2023) continues with the taxonomic reviews of capstone courses, focusing on the years 2007–2022. They noticed that external customer was often missing and that the courses were often shorter than one would expect for a realistic capstone experience. Recently, one can identify emphases on multidisciplinarity (Heikkinen and Isomöttönen, 2015), creativity (Apiola et al., 2012; Isomöttönen et al., 2019), and open-ended problems (Daniels, 2011; Isomöttönen et al., 2019). Perrenet et al. (2000) concluded that realistic project work (in place of problem-based learning) is good pedagogy for later stages of engineering curricula because students are then prepared for open-ended and complex problems.

2.3. Studies focusing on group work

This section raises a few examples of project education or otherwise relevant group work studies with the intention to raise interest toward the present mapping study. For instance, the examples below convey that students tend to face difficulties in their group work, which makes it interesting to look at the foci and theoretical frameworks in our included primary studies. It is equally interesting to look at how the pedagogic promise of group work, narrated in several project course reports, appears by means of our secondary research.

Iacob and Faily (2019) reported that students' initial positive expectations of group work did not often materialize in software engineering projects, including disappointment regarding commitment in groups. Similarly, Isomöttönen (2011) conceptualized group work as a trade-off because available positive experiences could easily turn into an observation of not all peers contributing or personally expected learning possibilities being constrained in the group. A consequence is that students can develop negative attitudes toward group work (Livingstone and Lynch, 2000; Wong et al., 2022), and prefer self-study (Wong et al., 2022). Previous negative and positive group experiences within a cohort shape the students' actions if they are granted the freedom to select (Allan, 2016). On these premises, an unsurprising challenge documented in higher education studies is the students' fear of unpredictability (Hinds et al., 2000; Strauss et al., 2011; Yorra, 2012). This refers to the students' worries about how their peers will contribute in a newly formed group and how the efforts of the peers will influence their grading. Relatedly, Poort et al. (2022) reported from an intercultural setting that it is essential to attempt to build trust early on in teams.

Group processes have been analyzed through participation patterns or roles, illustrated in the papers by Pieterse and Thompson (2010) and Marshall et al. (2016). One problematic role documented in these and many other studies is social loafing, colloquially referred to as free riding. Social loafing was defined as a phenomenon in which an individual's effort decreases when a task is set as a group task in place of an individual task (Latane et al., 1979).

Moreover, it is not evident that educators choose to intervene in students' problematic group situations. Burdett (2007) reported that some of the educators opined that students themselves should resolve their issues. All study participants, regardless that some of them found early interventions by teachers important, emphasized student initiative with problematic group situations. This study also considered that group pedagogy may be initiated merely for resource pressures, as fewer markings need to be done when students are evaluated as groups. Altogether, Burdett's study reminds us that students' group work is influenced by how we educators relate to it. Burdett argued that educators should supportively live with their student groups, on which the present authors agree.

Isomöttönen et al. (2019) presented a critical investigation of how students responded to an open-ended project assignment. Findings indicated that personal relationships with creativity and education system-imposed pressures could complicate students' participation in group work. Additionally, when a group of students did not have similar interests regarding a self-ideated project topic, diversity in the group was regarded as a difficulty. Although this study did not set out to study group work, in particular, these findings convey that the students' group work is influenced by many factors of a rather different nature.

Intervention-like efforts to improve students' group work condition includes strategies for removing a non-contributing member (Oakley et al., 2004; Pieterse and Thompson, 2006), educating students about group work (Kamau and Spong, 2015; Fronza and Wang, 2017), and arranging discussions in which group situation is addressed in the presence of the whole group (Clear, 2002; Isomöttönen, 2014). Additionally, Marques and Ochoa (2014) gave students a documented problem pattern (so-called thinkLet) when they heard a potential group problem in a project meeting. The goal was to help the group address its problem independently with the help of the pattern documented as a thinkLet. Students' feedback of thinkLets was mostly positive.

Even without systematic searches, researchers' interest in group work in relation to assessment is noticeable. A chief measure has been various peer- and self-assessment techniques enacted during group activities, with the concurrent assumption of supporting students' learning about group work (e.g., Wilkins and Lawhead, 2000; Hayes et al., 2003; Smith and Smarkusky, 2005; Clark et al., 2005; Farrell et al., 2012; Bastarrica et al., 2019). These actions arise from the need to assess group work fairly and have much focused on the identification of individuals' contributions to group work. Another area that has received ample attention is group formation, which is demonstrated by several secondary studies cited in the next section. One summarization in this area is that different group formation strategies introduce different pros and cons (Fincher et al., 2001).

In the main, group work is favorably referred to in conjunction with student-centered learning in project education reports (e.g., Souza et al., 2019; Pérez and Rubio, 2020). Students' feedback indicates that they enjoy and benefit from a project-based peer learning setting (e.g., Ahmad et al., 2014). When things go well in a group in challenging projects, students state that they value the whole group commitment (Isomöttönen et al., 2019) and identify a together-we-survive spirit (Isomöttönen, 2011).

Additionally, interesting examples are found outside the present educational or disciplinary context. For instance, the above-described group pedagogy as a double-edged sword (anxiety vs. valuable peer learning) shows similarly in studies on CS collaborative learning activities, without the emphasis on projects (e.g., Falkner et al., 2013). The study by Theobald et al. (2017), conducted in the context of biology, in turn, indicated that comfort experienced regarding being in the group influenced students' learning results. Better results were observed when students were given turns in collaborative activities in place of one student dominating the work.

2.4. Previous systematic literature studies

We identified several literature studies that addressed group work, and in the following paragraphs, we will focus on those published in the past ten years (i.e., from 2010). This time frame aligns with the period of our mapping study, the review of which can demonstrate a research gap. Notably, Petersen et al.'s mapping guidelines were published in 2008, just before this period. Before proceeding, we want to highlight an older review by Helle and Tynjälä (2006) because their observations comply with ours: Helle and Tynjälä (2006) observed that a majority of their 22 included papers were course descriptions without a more profound interest in evaluating or proving the effects of a course. The review was not limited to the present computing field but this observation appears similar to what we made during our mapping process: course description (experience) papers are well presented in our included papers, and, in particular, a large number of excluded papers were observed to be course descriptions, excluded for not being focused on group phenomena.

Henttonen (2010) reviewed research on social networks in small teams with an aim to clarify if the networks had influenced team effectiveness (32 included papers). The review was not restricted to an educational or another specific context. The density of communication had received ample attention in different kinds of teams and was found to have a positive influence on student teams in particular. Social networks' influence on behavioral and attitudinal aspects was observed to lack research results compared to the aspects of objective and perceptual performance. A key conclusion was that more research is needed to develop explicit causal links between social networks and team effectiveness.

Dugan (2011) focused on the issues of CS project courses at the level of whole courses and projects (with a survey of ~200 papers[1]). Due to this focus, their study adds to the taxonomic works cited above. The author observed a lack of attention to explicit links between learning theories and projects, and that instructors had struggled with how to support students' writing skills needed in producing project deliverables. Group work was observed to be challenging due to CS students being introverts and instructors lacking experience in teaching group work. No evidence that would have supported a particular grouping strategy was found, and altogether the group-work-related conclusion was that the instructors should pay attention to group selection, group size, and group organization.

Riebe et al. (2016) performed a review on the use of teamwork pedagogy in economy higher education (57 included papers). With this focus, the number of included papers is rather low. Due to the higher education context, the study can nevertheless inform the computing field. The most prevalent pedagogical themes in their sample were skills training, collaborative/cooperative experiential learning, and self and peer assessment. The study also revealed an emphasis on the topic of social loafing in groups, and so in connection with distributive justice experienced by students. Additionally, Tuckman's theory had received ample attention. The authors concluded a need for focused research that would increase understanding of whether the use of group pedagogy pays off from the perspective of students' gains, educators' efforts, and institutions' attempts to meet employers' expectations.

McEwan et al. (2017) conducted a review and a meta-analysis on controlled experiments on group work interventions that sought to enhance group efficiency and performance (51 included papers). Of the different intervention approaches, significant effects were found for workshop training, simulation-based training, and team reviews, while all training methods were concluded to improve group performance. The main takeaway was that group work training pays off. The study was not limited to either industrial or educational settings.

Also published in 2017, DeFranco and Laplante (2017) studied literature on software engineering teams' communication but neither this study was limited to an educational context (71 included papers). Communication had been commonly studied from the perspectives of global software engineering, effective teamwork, and project effectiveness. The authors further identified two key subthemes. The first related to global software engineering, and was an emphasis on tools and strategies for improving communication and team performance. The second related to effective teamwork and was ideal team dynamics and composition. An overall conclusion was the recognition of team effectiveness as a major challenge due to many aspects influencing it.

Taking a mapping approach, Costa et al. (2020) recently looked into group formation techniques in SE industry during the past ten years (51 included papers). The key observation was that computer algorithmic techniques had received the most attention. The topic of group formation appears to have received plenty of attention in computer-supported collaborative learning/work (CSCL/CSCW), therein mapped by Cruz and Isotani (2014) (48 included papers), Borges et al. (2018) (106 included papers), and Harris et al. (2019) (35 included papers). These papers were published between 2014–2019. In 2018, Reis et al. (2018) mapped affective and socio-emotional factors in CSCL (66 included papers) and identified group formation as a key area to consider in relation to affective factors as an outcome of their mapping. Although CSCL/CSCW research is concerned with group performance, the point of departure is the technology-enabled setting, which indicates less attention to contexts such as an authentic SE/CE project education setting. We included these papers in our mapping if they clearly demonstrated a CS/SE project-based learning setting.

Taken together, none of the secondary studies above clearly focused on mapping research on group phenomena in project-based learning in CS/SE educational context. The review studies did not (by definition) deliver a broad overview of their respective fields, and several of them were not devoted to an educational or the present disciplinary context. The available mapping studies focused on group formation or another specific area of group work. It is concluded that the present effort can fill the gap still observable in secondary research regarding group work in the tradition of project-based learning in CS and SE fields.

3. Research method

3.1. Selection of the research target

Three main reasons for selecting the research target can be summarized: the significance of the nexus between project-based education and group work skills, the historical focus on course descriptions and experience reporting, and the lack of a big-picture study that could advance the field.

Regarding the first reason, the multiple taxonomic works raised in Section 2.2 underscore projects as flagship education in computing, while our Introduction and Section 2.3 outlined the importance of enhancing graduates' group work competencies. Generally, the need to expose students to the realism of working life appears to be constantly acknowledged (see, Section 2.1).

Regarding the second reason, historically, research has often taken the form of a course description accompanied by experiences. This is indicated, for instance, by the results of Helle and

[1] The approximated number of papers might not be comparable with the other secondary studies presented in this section: Dugan (2011) does not specify the exact number of the primary studies and the method of the study is not described comprehensively.

Tynjälä (2006). Similarly, Tomayko (1998) noted that new project course reports are submitted because isolated teachers are excited by the authentic projects as something new. Additionally, papers reporting case implementations of project- and problem-based learning were frequently pinpointed during the present mapping.

The third reason follows from the first two. Given this appearance of the research tradition, an attempt to disclose the big picture of the research on group work in project courses was considered necessary to advance the research field.

Systematic mapping studies make sense of the research fields as a whole whereas systematic literature reviews target at providing synthesis on a specific research question (Kitchenham and Charters, 2007; Petersen et al., 2008; Kitchenham et al., 2011). This difference shows in the results of respective studies: presenting high-level maps versus synthesizing answers to specific questions—mapping studies do not primarily target aggregating the results of the original papers (Kitchenham et al., 2011). In the former, research effort is expended on categorizing a high number of papers. In the latter, the substantive effort is rather the synthesis. Mapping studies incorporate both general questions (e.g., types of publications and research methods used) and domain-specific questions (e.g., what theories have been referred to in the field) (Petersen et al., 2015). The benefits of systematic mapping include that they help identify foci and gaps in the literature, and thus direct the research in their respective field (Petersen et al., 2008). Mapping results also guide subsequent, narrower review studies (Petersen et al., 2008; Kitchenham et al., 2011). The present research interest encouraged a mapping study in place of a review. The guidelines proposed by Petersen et al. (2008, 2015) were followed.

An interesting methodological choice could have been the snowballing literature study proposed by Wohlin (2014). This method begins from a starting set of papers and proceeds by both backward and forward snowballing and hence emphasizes a well-reasoned selection of the starting set. Thus, the method is not initially based on database searches as in the present study. Because of the present holistic interest in the research field, mapping based on database searches was commenced. The authors opine that the snowballing approach provides an option for follow-up review studies. With the help of the starting set approach, the reviews could potentially extend to studies published earlier than the present timeline without introducing an unmanageable amount of work.

3.2. The protocol

This section describes how the method was applied. The research group consisted of the first (MK) and the second (VI) author.

3.2.1. Research questions

To form an overview of the research on group phenomena in project-based learning published in the past decade, we used the following mapping questions:

(Q1) When and in which forums has the topic been researched?
(Q2) Where the topic has been researched geographically?
(Q3) What types of studies have been favored in the research area?
(Q4) What kind of research methods have been employed in the studies?
(Q5) What areas of group work the studies have been focused on?
(Q6) What theoretical frameworks have been used?
(Q7) What has been the context of the education in the studies?

Q1, Q2, and Q4 clarify where and how the research has been conducted. Q3 was answered by classifying the studies by their type according to the classification system provided by Wieringa et al. (2006). The rest of the questions are topic-dependent and concerned with the focus of the research area. To answer the last question we classified the papers into two categories based on their focus on either undergraduate or graduate studies.

3.2.2. Data sources and search strategy

We selected ten online databases,[2] including the (1) ACM Digital Library, (2) IEEEXplore, (3) Taylor & Francis Online, (4) Science Direct, (5) Springer Link, (6) Scopus, (7) ISI Web of Science, (8) ProQuest, (9) Wiley Online Library, and (10) Google Scholar. We did not conduct manual searches for scientific journals since we observed that papers from relevant journals had been indexed for the past decade in Science Direct, Scopus, or Taylor & Francis Online, and thus already captured by our selection of databases.

We selected databases with the aim of covering central publication channels and ensuring that relevant data was not missed. However, the methodological literature (Petersen et al., 2008, 2015) does not take a clear position on the selection of databases and delimitation of data. Our selection was influenced by initial exploratory searches we conducted to develop search phrases and evaluate the scope of the mapping. During this stage, we inspected Google Scholar hits to determine the databases that produced relevant hits. We selected the databases[3] based on their prevalence in these hits, as illustrated in Fig. 1. The figure shows the number of Google Scholar search results for each WWW domain name in a particular exploratory search, ordered by the number of hits per domain. The steep decline in search results per domain name suggests that the "excluded tail" does not include central forums for our topic.

The flexibility and intuitiveness of the user interfaces of some selected databases have previously been questioned (Dybå et al., 2007), and our experiences with initial exploratory searches 13 years later still support this. As a result, we used Google Scholar as an extra safeguard to capture any relevant results that may have been missed from searches made using the databases' own user interfaces. To do this, we first captured the entire Google Scholar result set and then filtered out links that did not point to the selected databases.

The development of the search strategy took three weeks and was carried out iteratively. In each iteration, we collected the search results corresponding to the strategy from all selected databases programmatically into a Sqlite3 database. We then queried this database with SQL to examine the quality of the search strategy. The authors met approximately twice a week to make improvements to the search strategy. Here, we employed a test–retest method and searched for synonyms of our search terms using the ERIC thesaurus descriptor search.[4] We also used the keywords from the papers captured in our initial exploratory searches to refine our search phrases.

Table 1 presents search phrases grouped according to concepts "project", "group work", and "CS/SE", which correspond to the aim of our study. The same presentational convention for search phrases was used by Fernandez et al. (2011). In Table 1, search terms that belong to the same thematic concept (such as "project") are interchangeable and were combined using the Boolean OR operator. We used the Boolean AND operator between concepts to ensure that the search phrase always includes at least one search term from each thematic concept.

2 For simplicity, we use the term database to refer to all of our data sources, regardless of whether they are databases or search engines.

3 Scopus and ISI Web of Science are search engines in nature and therefore did not appear in exploratory searches.

4 Available online: https://eric.ed.gov.

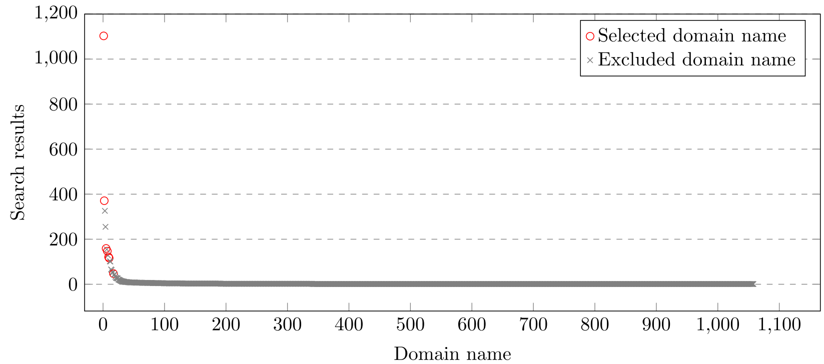

Fig. 1. The number of Google Scholar results per domain name of the link. The two excluded domains with the high number of hits are researchgate.net and google.com. Most of the domains have only one hit.

Table 1
The search phrases. Terms with "*" can be either singular or plural. The regular expression "(-|␣)" between terms indicates that terms can occur either alone or together combined with a hyphen.

Concept	Search terms and synonyms	
Projects	(''project(-|␣)based learning'' OR pbl OR capstone OR ''student project*'' OR ''team project*'' OR ''group project*'' OR ''problem(-|␣)based learning'')	AND
Group work	(''group work'' OR ''team work'' OR teamwork)	AND
CS/SE	(computing OR ''computer science'' OR ''software engineering'')	

To give priority to the "project" concept in the search results, we utilized the title, abstract, and keyword search fields available in IEEEXplore, Scopus, and ProQuest. However, in other databases, we had to employ all meta fields and full text, which led to an increase in irrelevant search results. In contrast, the concepts "group work" and "CS/SE" was given less emphasis, and as a result, we used all meta fields and full text for these concepts. To ensure compatibility with each database's specific format, we tailored the syntax of the search phrase accordingly, as the syntax for defining search term fields varied among databases.

The final searches were conducted between the 14th and 15th of May 2020 and extended to cover the research until the end of 2021 on the 6th of March 2022. We programmatically collected the search results and stored them in a Sqlite3 database, including the title, URL, DOI, abstract, and a combined meta field that contained author information, forum, and year of publication. If a search interface allowed us to download results as a CSV file, we used a node.js script to save the corresponding rows into our database. Otherwise, we used a bot with the Puppeteer library to collect the results by navigating the search interfaces like a regular user. To begin the search process, the bot launched the Google Chrome web browser and directed it to the appropriate search interface (such as https://dl.acm.org). The bot then executed the search query using the parameters specified in Table 1, read the search results displayed on the page, and stored them in our results database. The bot continued to the next page of results, repeating this process until it reached the last page. If the search interface did not provide a full abstract of the result on the results page, the bot would navigate to the corresponding URL and extract the abstract, which was then added to our results database. The bot was deliberately slowed down to prevent unnecessary load on the search interface and to enable real-time monitoring via computer screen. To ensure the accuracy of the data, we verified the correctness of the saved rows for each database by thoroughly inspecting small samples of rows and ensuring that every field was valid. Additionally, we visually scanned the entire search results to identify any incorrect rows. The programmatic collection method was used to avoid human error and reduce workload. Furthermore, it enabled us to perform searches all at once and store the results immediately in our result database.

The search results per database are shown in Table 2. The searches were performed in the same order they are presented in the table. Thus, the number of duplicates in Table 2 refer to those papers that were already in our database when the search corresponding to the row was conducted. It should be noted that Table 2 has the combined results of two separate searches. In 2022 we extended the original search performed in 2020 to cover literature until the end of 2021 and the extension was performed in the same order as the original search. The original search and the extension slightly overlap in the year 2020. This is particularly noticeable for the first database in the list, ACM DL, which had 49 duplicates that would not have been there without the overlap. It even had 5 duplicates in the original search, which resulted from a technical issue in the search interface. The order of the results changed slightly when moving between pages, causing some results to appear twice and others to be missing completely. The 5 missing results were added manually to our results database. We selected eligible papers from the unique search results remaining after removing duplicates. After the paper selection, we conducted a backward snowballing search, in which the references of the selected papers were screened. This was done first for the papers selected from the original search and later for the search extension. The number of relevant papers found through snowballing is also presented in Table 2

As shown in Table 2, out of 4972 Google Scholar search results 2939 were also found in the other databases, which is 74.92% of all unique search results of the other databases. This set includes the majority (67 of a total 106) of the "selected" search results that appeared in the Google Scholar search, but still, 39 eligible papers were found only in the Google Scholar's unique search results. The use of Google Scholar as an additional verification to ensure efficient search thus seemed to be beneficial although the selection rate (1.9%) was lower than the average of the other databases (4.2%). The lower rate implies that, although Google Scholar found about half (50.5%) of the relevant search results, it also found more irrelevant results than the other databases and search engines we used.

3.2.3. *Selection criteria*

Papers that met all of the following inclusion criteria were included in the study:

1. the paper focuses on a phenomenon or aspect of group work,

Table 2
Search results in the selected databases. The numbers in parentheses indicate the growth of the search results in the search extension from the middle of May 2020 to the end of 2021, and are included in the totals.

Source	Total	Duplicates	Unique	Selected
ACM DL	1300 (+308)	49 (+44)	1251 (+264)	61 (+10)
IEEExplore	653 (+160)	59 (+28)	594 (+132)	50 (+12)
Taylor & Francis Online	186 (+37)	23 (+20)	163 (+17)	5 (+0)
Science Direct	758 (+163)	21 (+20)	737 (+143)	21 (+0)
Springer Link	105 (+9)	6 (+5)	99 (+4)	0 (+0)
Scopus	1129 (+330)	665 (+170)	464 (+160)	19 (+9)
ISI Web of Science	48 (+15)	32 (+9)	16 (+6)	2 (+0)
ProQuest	296 (+29)	52 (+14)	244 (+15)	3 (+0)
Wiley OL	396 (+118)	41 (+11)	355 (+87)	3 (+0)
Google Scholar	4972 (+1335)	2939 (+735)	2033 (+600)	39 (+8)
Total before snowballing	9843 (+2504)	3887 (+1056)	5956 (+1448)	203 (+45)
Backward snowballing	2097 (+241)	539 (+122)	1559 (+120)	22 (+3)
Total	11 940 (+2745)	4426 (+1178)	7515 (+1567)	225 (+48)

   (a) e.g., assessment of group work, teacher's perspective on improving group work, online group work, group dynamics, experiences of group work, pair work, or soft skills clearly related to group work.

2. the paper focuses on project-based learning
   (a) We used classification by Morgan (1983) (see, Section 2.1) to include paper if we interpreted authors reporting on a "project component" rather than a single "project exercise" as a minor part of a course.
   (b) e.g., project assignment that was interpreted to indicate complex work over a longer period than a weekly CS exercise.

However, if a paper that met the acceptance criteria met any of the following exclusion criteria, it was excluded from the study:

1. the context of the paper is not tertiary education,
2. the paper is published before 2010,
3. the publication is a book, dissertation, abstract, poster, "work-in-progress" paper or a brochure of a panel discussion,
4. the paper is not a primary study,
5. it is not clear from the paper whether it is in the context of CS/SE teaching,
6. the paper is not written in English,
7. the paper has not been convincingly peer-reviewed,
   (a) e.g., the publication is at level 0 of Publication Forum classification system of Finland.[5] This is the key quality assessment activity. Each paper that fulfills the other inclusion criteria is manually checked for its Publication Forum level,
   (b) e.g., the description of the peer review process of the publication is not available with reasonable effort or it is incomplete or superficial. This is inspected if the publication does not have a Publication Forum level indicated.
8. the research interest or viewpoint of the paper is not in the context of teaching,
   (a) e.g., the paper focuses on a group work phenomenon in an industrial or context-free environment, even if it was a case study in the context of a CS/SE course.
9. it is not clear from the paper whether group work is studied in a project environment,
10. the group work's project context is not related to any activity related to a software engineering project,
   (a) e.g., a CS or SE course is a research project, research ethics project, or another general topic not relatable to SE project activities.
11. the context of the paper is related to group work, but the research viewpoint or interest is not
   (a) e.g., if the topic is related to the tooling of collaboration, the focus is purely on the development of the tool and not on the relevance of the tool in the context of group work.
   (b) e.g., the tool being reported does not have a meaningful relation to a project setting.

3.2.4. Selection of the primary studies

The selection process is illustrated in Fig. 2 and its application in Table 3. The process can be divided into two phases: (1) establishing the selection criteria and (2) applying the criteria to the data. In Fig. 2 phase 1 is represented by steps 1–3. In step 1 the initial version of the selection criteria was established. The criteria were used to classify the papers either as "approved" or "rejected" based on whether they passed the selection criteria when inspecting the paper title, abstract, and keywords. If these did not provide enough information to make the decision or the paper was a borderline case, it was classified as "uncertain".

Steps 2 and 3 were used iteratively to ensure that the selection criteria were interpreted and applied to the data in a consistent manner. A consistent interpretation practice was developed using Cohen's Kappa value. Steps 2 and 3 were repeated until Cohen's Kappa reached an acceptable level. The indicator was calculated by comparing the decisions between the raters for a certain amount of papers. Initially, we used a set of 50 papers per calculation. The acceptable level for Cohen's Kappa is 0.6 or higher, but the 95% confidence interval for the indicator with our initial setting of 50 rating decisions was unacceptably large (between 0.25 and 0.83). As a result, the number of ratings used to calculate Kappa was later increased to 100 and then again to 150. Kappa was calculated only on "approved" and "rejected" categories because the "uncertain" category was used at a low threshold to postpone the classification decision. It is worth mentioning that Kappa was considered only as a tool to test the consensus of the authors on classification decisions. The real goal was to find the actual consensus.

The latter phase of the process consisted of steps 4–10, where the selection criteria were applied to the data. The work was divided between the authors so that VI classified the first 800

[5] Available at: https://julkaisufoorumi.fi/en/evaluations/classification-criteria. Level 0 means that the publication does not meet the level 1 criteria, which among other quality criteria includes a plausible peer-review process. Other quality criteria contain registered ISSN or ISBN number, sufficient transparency, scientific focus, editorial board, scope, and credibility of the publication channel.

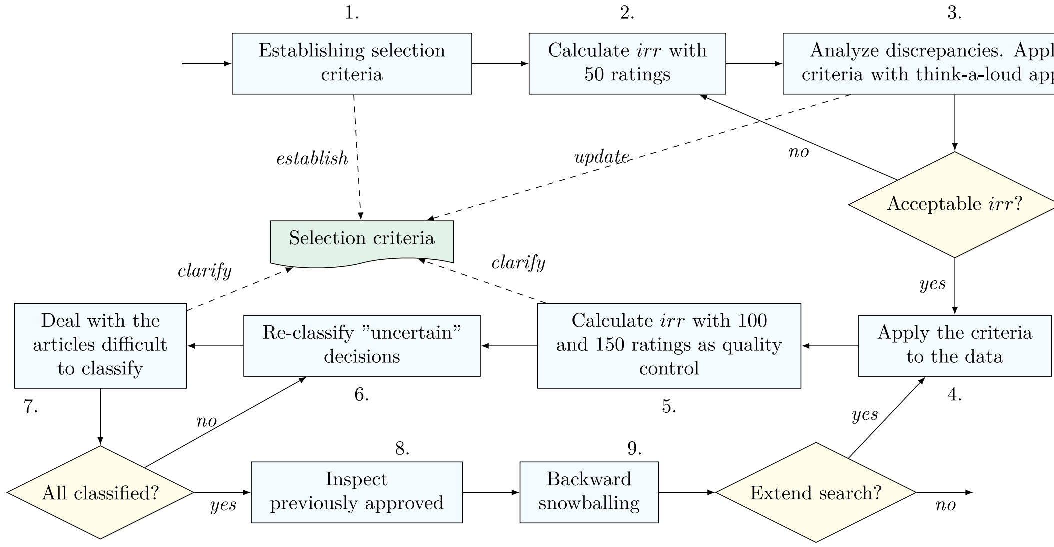

Fig. 2. Selection process. *irr* (inter-rater reliability) refers to Cohen's Kappa. The flow of the selection process is represented by solid lines, while dashed lines indicate when and how the selection criteria were modified. Activities are denoted in blue, decision points in yellow, and the "Selection criteria" entity in green. The log of the selection process is presented in Table 3, which documents the steps taken during the process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and MK the following 1000 papers. As a further quality control, a random sample of 50 papers was picked from both sets to be used in a calculation of Cohen's Kappa, so that a total of 100 papers were evaluated by both authors and used in the calculation. The calculated Kappa was 1, i.e., the selection criteria were applied in a completely uniform manner. The remaining papers were classified by MK. In Steps 6 and 7, a total of 403 papers left in the "uncertain" state were re-classified either to the "approved" or "rejected" category based on the full text of the paper. The papers that were difficult to classify were discussed together, refining selection criteria as necessary. Once all of the "uncertain" papers were re-classified, the accepted papers were re-evaluated because the selection criteria had become slightly stricter than they were at the beginning of the process (Step 8).

In Step 9, we performed a backward snowballing. This was done by first collecting all the references of the included papers and then filtering irrelevant references out. Our decision to store the search results – including the web URLs – in a database proved to be especially useful at this phase, allowing us to collect most of the snowballing data automatically with a web scraper. Still, a relatively big part of the data was collected manually from the PDF files or databases. Then, the selection criteria were applied to the snowballing data in the same manner as for the initial data set.

We encountered 13 cases where the Publication Forum level marking was not identifiable. To evaluate these cases, we conducted collaborative research sessions and assessed each one in video meetings against the criteria for Publication Forum level 1, as well as other inclusion criteria mentioned in exclusion criteria 7b. As a result, we rejected 8 cases. This thorough process took place from January 6th to March 25th, 2021.

After the initial snowballing step in 2020, we extended the search in 2022 to include literature until the end of 2021 and performed steps 4, 6, and 7 for the new data. We considered step 5 unnecessary because the selection criteria were already established and the consensus of the authors on executing it had been tested. Also, step 8 was not needed for the new data, because the selection criteria did not evolve in the process.

3.2.5. Data extraction

To extract the desired data from the selected papers, keywords were constructed for each paper so that the composition of the keywords could be used to answer the research questions. Thus, the mapping is completely based on keywording the data. The keywords were stored in a Sqlite3 database, allowing them to be processed efficiently with SQL statements and thus perform the data extraction step with a reasonable amount of work. A custom-made user interface was used to set the keywords.

For Q1, the publication year of the study and the name and abbreviation of the forum were keyworded. If the forum did not have an official abbreviation, an abbreviation according to the ISO-4 standard[6] was used. As for Q2, the keywording was done with the locations marked for each author of the study. With regard to Q3, the classification system provided by Wieringa et al. (2006) was used. For Q4, Q5, and Q6 keywords were analyzed without a pre-selected taxonomy. The researcher's judgment of classifying the papers into pre-selected categories would have raised a need for additional validity measures (e.g. calculating Cohen's Kappa) to test the consensus between the researchers. To answer these questions, only the relevant parts of the paper were read, which for Q4 mainly meant a section describing the research method, for Q5 either the summary or the section describing the results, and for Q6 the research method or analysis parts of the study. For Q7 the keywords used were "undergraduate" and "graduate".

With questions Q4–Q6, the keywords with the conceptually same meaning were combined to improve their statistical quality. Because all papers were not read in their entirety, the selected papers went through a validation process, in which their full text was programmatically searched for each keyword defined to find possibly missed keywords. In the same process, each keyword was re-evaluated and changed or removed if it was considered inaccurate. Because the majority of the keywording was initially done by the first author, a random sample of 20 selected papers was additionally independently keyworded by the second author. The keywords that caused discrepancies were then either changed, removed, or left as is for each paper they were attached to.

6 Available online: https://www.iso.org/standard/3569.html.

Table 3
Applying the selection process.

Step	Description	Date	Notes
1	Establish selection criteria	4 May 2020	Selection criteria is updated several times afterwards
2, 3	Calculate *irr* with N = 50 − 24[a] and analyze discrepancies	20 May 2020	Kappa: 0.54, [0.25, 0.83][b], agreement: 77%.
2, 3	Calculate *irr* with N = 50 − 12[a] and analyze discrepancies	22 May 2020	Kappa: 0.80, [0.54, 1][b], agreement: 94.7%.
2, 3	Calculate *irr* with N = 100 − 15[a] and analyze discrepancies	25 May 2020	Kappa: 0.67, [0.4, 0.94][b], agreement: 93.6%.
2, 3	Calculate *irr* with N = 150 − 27[a] and analyze discrepancies	26 May 2020	Kappa: 0.72, [0.52, 0.93][b], agreement: 95%.
4	Apply the selection protocol	27 May–21 June 2020	VI 800 pcs, MK 1000 pcs
5	Calculate *irr* with N = 100 − 21[a]	22 June 2020	Kappa: 1, [1, 1][b], agreement: 100%.
4	Apply the protocol	23 June–31 July 2020	Rest of the unclassified. Mostly by MK
6	Re-classify "uncertain" decisions	3 Aug–7 Aug 2020	VI & MK
7	Dealing with the papers difficult to classify	7 Aug 2020	The criteria were clarified
6	Re-classify "uncertain" decisions	7 Aug–10 Aug 2020	VI & MK
7	Dealing with the papers difficult to classify	10 Aug 2020	The criteria were clarified
8	Check previously approved	17 Aug–21 Aug 2020	Because the selection criteria was changed
9	Backward snowballing	12 Oct–28 Dec 2020	Applied steps 4, 6, 7, 8. Discussed discrepancies.
	Quality control of selected papers	6 Jan–25 Mar 2021	Determined quality for forums not in the Publication Forum classification system of Finland.
Selection process for the extension to cover the literature until the end of 2021:			
4	Apply the selection protocol	8 Apr–1 Aug 2022	MK. "Uncertain" was used at a low threshold
6	Re-classify "uncertain" decisions	1 Aug–15 Aug 2022	VI & MK
7	Dealing with the papers difficult to classify	17 Aug 2022	VI & MK. The criteria did not change
9	Backward snowballing	16 Nov–21 Nov 2022	Applied steps 4, 6, 7. No need for 8 because of unchanged criteria

[a]The value of N follows the format of $N =$ *(all ratings) − ("uncertain" ratings).*
[b]95% confidence interval is reported in the brackets after the Kappa value.

3.3. Validity

Petersen and Gencel (2013) recommend taking the following validity attributes into account: descriptive validity, theoretical validity, generalizability, and interpretive validity.

3.3.1. Descriptive validity

Descriptive validity refers to the objectivity and accuracy of data gathering, ergo, the researchers' ability to record the data accurately. To reduce threats to descriptive validity, we focused on the quality of selection criteria and used a well-defined classification for determining the research type for the eligible papers. The need for researcher interpretation for defining the keywords for the papers was considered relatively small, yet we did several validation checks to reduce the validity threat as described in 3.2.5. Also, the keywording was performed using self-developed computer software, which acted as an effective data extraction form. The threat to descriptive validity is therefore seen to be taken into account.

3.3.2. Theoretical validity

Theoretical validity is concerned with researchers' ability to capture what is intended. This may for example be affected by a biased selection of papers. All the selected papers were inspected multiple times because they were first screened based on title, abstract, and author-set keywords, later read for keywording and re-evaluated for inclusion in Step 8 of the selection process (See., Fig. 2). Also, the uniform interpretation of the selection criteria between authors was tested with Cohen's Kappa, as described in Section 3.2.4. Several actions were taken to mitigate the threat of missing relevant research in the search phase. The search phrases were developed in a rigorous manner as described in Section 3.2.2 and we used all the databases considered relevant (a total of 10) to reduce the risk of missing relevant research. The use of Google Scholar as an additional verification for efficient search, saving the whole set of search results to our result database in one step, backward snowballing, and the use of computer scripts to collect the search results are all measures to reduce the validity risk.

The majority of the keywording was initially done by the first author, which is the most apparent threat to theoretical validity. To mitigate this threat, a random sample of 20 papers was independently keyworded by the second author as described in 3.2.5. In addition, the selected papers went through a programmatically assisted validation step to find possibly missed keywords, which included re-evaluating every initially set keyword (see 3.2.5). We consider the performed validity measures on keywording at least comparable to other advanced systematic mapping studies (e.g., Petersen et al., 2015).

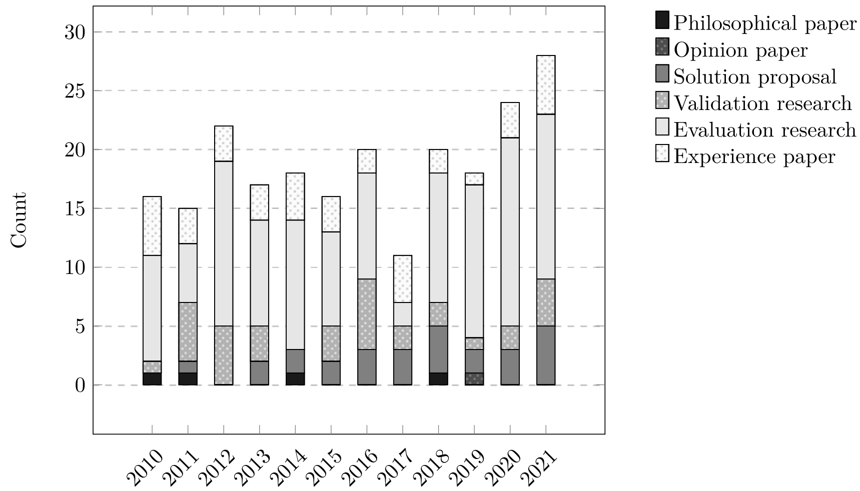

Fig. 3. The number of papers per year and the type of research.

3.3.3. Generalizability & repeatability

The study deals with project-based learning in CS and SE disciplines and cannot be generalized to project-based learning in other disciplines. However, the results can be useful when comparing another discipline to CS/SE. We followed the methodological guidelines of Petersen et al. (2008, 2015) to improve repeatability. The detailed reporting of the method makes it possible to repeat the study with a different research group. Yet, we acknowledge that there is some interpretation in the selection and classification of the papers, and another group of researchers may not end up with exactly the same results as we did.

3.3.4. Interpretive validity

Interpretative validity is achieved when the risk of researcher bias is mitigated. The first author is not an author in any of the selected papers, but the second author is in two. The risk was acknowledged and the referred papers were independently included by the first author, thus we do not consider this to be a bias in interpretation.

4. Results

In this section, systematic maps are used to answer the mapping questions presented in Section 3.2.1. The base data for the results is available online (Kokkoniemi and Isomöttönen, 2023).

4.1. Publication time and type

The papers appear to be distributed quite evenly for each year of the time span except for the year 2017 when only eleven papers were published. In Fig. 3, the yearly counts are presented as a bar graph by stacking the bars representing the types of papers on top of each other, so that the diagram also shows how the annual counts are distributed according to the types presented by Wieringa et al. (2006). Each paper is classified only according to the most representative type of paper. The graph implies a slight increase in the number of experience papers in 2020 and 2021 compared to previous years.

The ratios of the types of selected papers are illustrated in Fig. 4. Evaluation and validation research plays a central role in the selected papers, although the difference between the two types is not obvious, especially for case studies. Petersen et al. (2015) classify a case study as a validation study if it is conducted using students as subjects. They argue that, in turn, research conducted in an authentic "real-life" environment is evaluation research. However, as our focus is on the teaching environment, using students as subjects does not mean the environment would not be authentic. So, we classify a case study as validation research if it clearly focuses on piloting or testing some new solution, idea, method, or tool. Otherwise, the case study is classified as evaluation research.

A mentionable number of experience papers passed the selection criteria. These papers focused in some way on describing group phenomena in project education. The types of rejected papers were not recorded, but based on the observations made when applying the selection criteria, a large proportion of them appeared to be experience papers focusing on describing the implementation of a single course. Four philosophical papers and one opinion paper were also found, all of which also had features of the other types of research. Solution proposals could include small-scale testing of the proposed solution, for example with a questionnaire, but not convincing validation.

4.2. Publication fora

The yearly number of papers published in different forums is summarized in Fig. 5. The official abbreviation of the conference or journal is used if one exists. Otherwise, the abbreviation according to the ISO-4 standard is used. The majority of the publications are distributed quite evenly across 114 different forums, with most forums having 1–3 papers. However, there is a noticeable concentration of publications in the *Frontiers in Education (FIE)* conference, where a total of 25 papers have been published. The conference appears to have developed a tradition of discussing group phenomena in CS/SE project education.

Other popular conference proceedings include *ACM Special Interest Group on Computer Science Education (SIGCSE)*, *The Innovation and Technology in Computer Science Education (ITICSE)*, and *Conference on Software Engineering Education and Training (CSEE & T)*. Each of these conferences has published 7–14 papers addressing the topic during the last decade. In established journal publications the topic has been covered in *ACM Transactions on Computing Education* (6 papers), *Computers in Human Behavior* (4 papers), *Computer & Education* (4 papers), *IEEE Transactions on Education* (6 papers), and *Journal of Systems and Software* (6 papers).

4.3. Geographical distribution

The geographical location of the authors shown in Fig. 6 illustrates where the research on the topic concentrates geographically. In the figure, a single paper may be located in more than

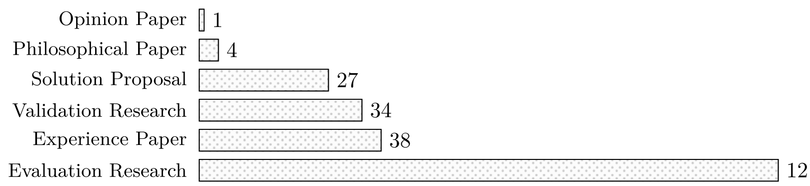

Fig. 4. The distribution of the papers according to the type of research.

Table 4
Geographical differences by the focus of the study in either undergraduate or graduate studies. A single paper could have been done in a collaboration between two continents, so the figures may include the same paper more than once.

Continent	Undergraduate		Graduate	
	Papers	Percent	Papers	Percent
Africa	2	100%		
Asia	20	90.1%	2	9.9%
Oceania	8	66.7%	4	33.3%
Europe	67	62.0%	41	38.0%
North America	53	72.6%	20	27.4%
South America	5	80%	1	20%

one country if it was done in a multinational collaboration, but the same country is counted only once per paper. Thus, the sum of the numbers presented in Fig. 6 exceeds the total number of the selected papers for the mapping.

During the last decade, most of the research on the topic has been conducted in the United States (73 hits). The difference to the UK (27 hits) in the second place by the publication count is remarkable. In both countries, most of the research is evaluation research, but compared to other countries, the number of experience papers appears to be high. Yet, the total number of papers for other countries may be rather small for meaningful comparisons by the type of research. Although the top of the list (in Fig. 6) is dominated by large Western countries such as Spain, Australia and Germany, smaller societies such as Sweden and Finland are also near the top.

Table 4 describes the distribution of research focus between different continents for either undergraduate or graduate studies. A single paper could have been done in a collaboration between multiple continents, it may focus on both undergraduate and graduate studies, or it cannot focus on either one of them. Thus, the numbers in the Table do not completely add up to the total number of the selected papers. In each continent, the research seems to focus more on undergraduate than graduate studies. In Asia, Europe, and North America, the volume of research (see., Fig. 6) gives reliability for the figures, whereas in Africa, Oceania, and South America the volume is relatively small, hence the numbers should be interpreted with caution. Europe and Oceania appear to differ from Asia and North America in terms of research focus, as the graduate studies attract relatively more research in Europe and Oceania than in Asia or North America. In Europe, 38% and in Oceania 33.3% of the research on the topic focuses on graduate studies, whereas for North America the same figure is 27.4% and for Asia 9.9%.

4.4. Research methods

Fig. 7 shows the prevalence of the research methods explicitly mentioned in the papers. The numbers are grouped by the type of research. Most of the research appeared to be empirical: either evaluation or validation research. A common example of such research was a case study in a context of an individual project course, which is then analyzed using qualitative-, quantitative- or mixed methods. Remarkably few randomized comparative experiments were used.

Interviews were used widely for data gathering. The interview could either be in-depth in nature (19 papers), more lightweight (20 papers), or a focus group session (8 papers). In Fig. 7, the different interview types are shown as separate bars. The total number of papers using some kind of interview as a data-gathering technique is 47, of which nearly all are classified as evaluation research.

Questionnaire surveys (74 papers) are mostly used to collect quantitative data, but some papers use them also to collect qualitative data. The different qualitative content analysis methods (20 papers) are combined together in the figure. Rather few papers explicitly mention the use of action research, ethnographic observation, or Grounded theory. This could be seen as indicative of a lack of approaches that are known to be challenging and require persistence or long-term commitment.

Based on our observations made during the selection of primary studies, a mentionable portion of papers focusing on project-based learning in CS/SE appeared to be experience papers without a clear research method. Although such papers clearly stood out, only 8 of these "Marco Polo" studies[7] passed the inclusion criteria of the selection phase.

4.5. Areas of group work

The distribution of research foci on different group work areas is depicted in Fig. 8. A single study may cover more than one area of group work, which explains why the numbers shown in the figure exceed the total number of papers. Four papers did not focus on any particular area of group work and are not included in the numbers. The terms used in Fig. 8 were collected from the selected papers and synonyms were combined together under one term. The interest in the research area was clearly focused on a few areas of group work. The assessment of group work appeared to be the most researched topic in the area of CS/SE project education research. It was addressed in a total of 48 papers, which were mostly evaluation research, but interestingly validation research, solution proposals, and experience papers appeared to be quite evenly represented with 7–10 papers each. While still included in the numbers of "assessment" (48 papers) the papers about peer assessment are separately presented in Fig. 8. Peer assessment was addressed in 24 papers, implying that half of the assessment research focused at least partly on peer assessment.

Also, group's communication appeared to be one of the most attractive areas for researchers (41 papers). The research on communication consisted mainly of evaluation research and experience papers. It appeared in seven papers together with a focus on group's collaboration, indicating a weak link between the two areas. A total of 35 papers focused on collaboration, representing each of Wieringa et al. (2006) types except opinion papers. Among these, there were even two philosophical papers, but a majority of the papers (19 papers) were evaluation research.

[7] We use the term Marco Polo as defined by Pears et al. (2005).

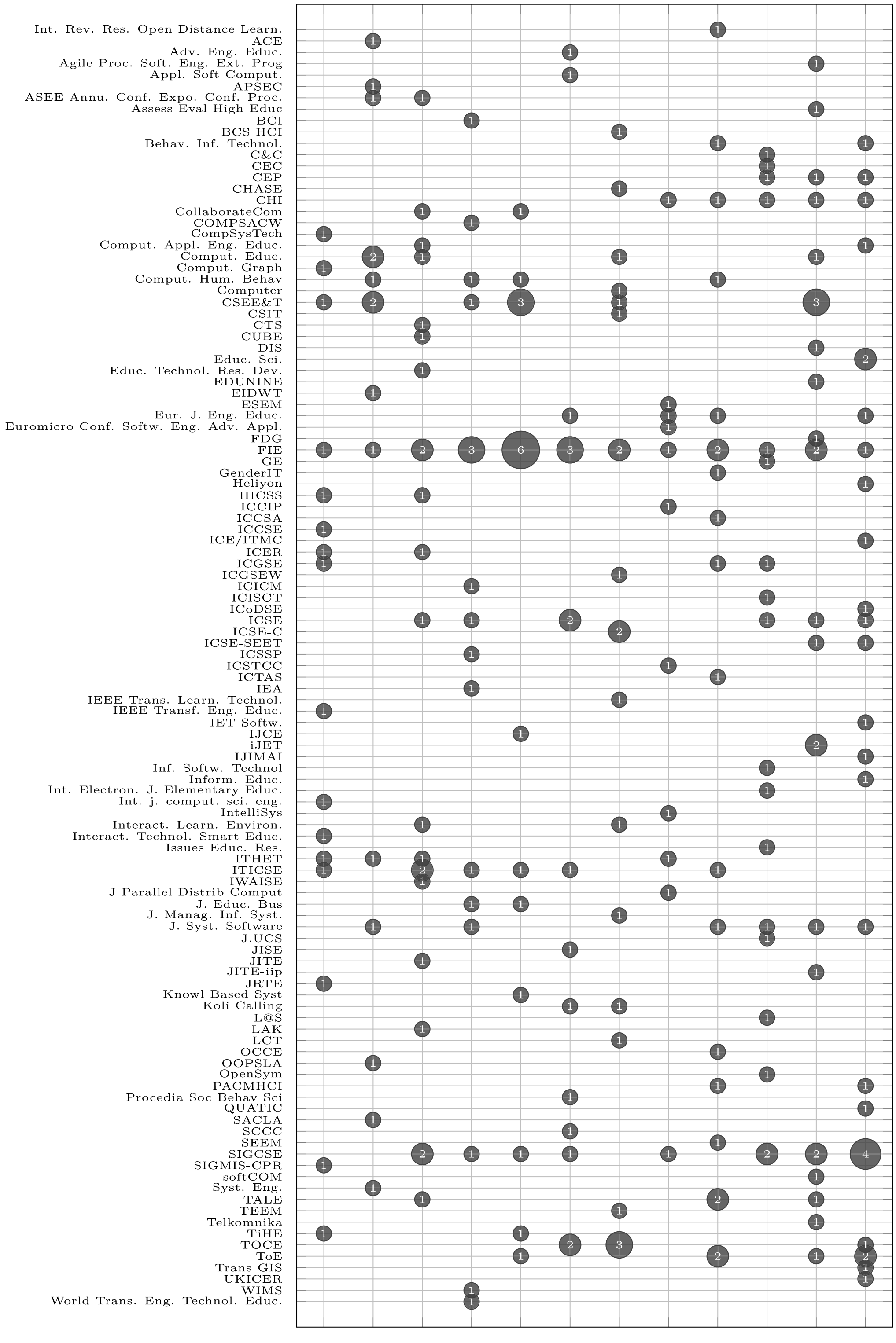

Fig. 5. Annual number of publications in different forums.

The research focusing on *group formation* also appears to play a key role (32 papers). The problem of successfully forming student teams concerns every educator supervising a group project course. These papers focused for example on investigating the most optimal way to form groups with various algorithms and heuristics (e.g., Jahanbakhsh et al., 2017; Nand et al., 2018), as well as validating the selected techniques. A closely related topic for group formation is the composition (or structure) of the team (12 papers). These papers were concerned about the attributes (e.g., role, age, demographic traits, size of the team, personalities) of which a good team consists (e.g., Pieterse and Thompson, 2010). A majority of *team composition* papers were evaluation research. In turn, although *Team building* (11 papers) is related to the formation and composition of a team, it focuses on the

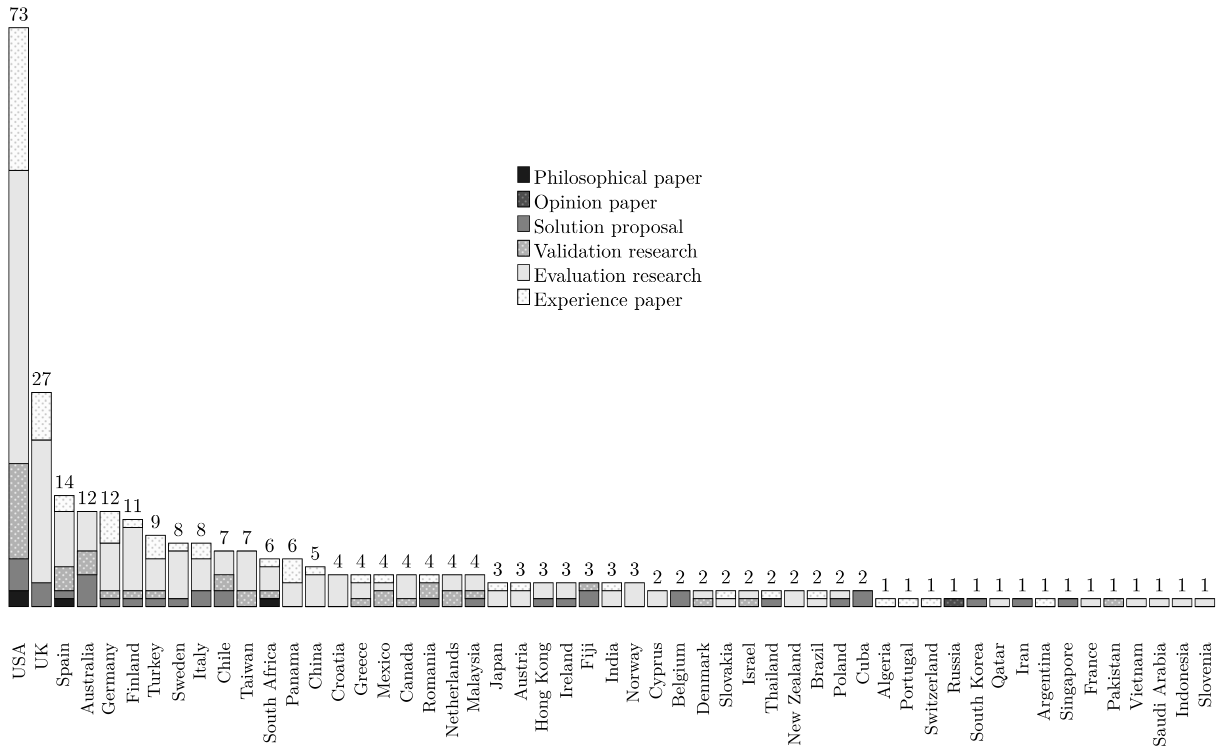

Fig. 6. Geographical distribution of papers and the type of research.

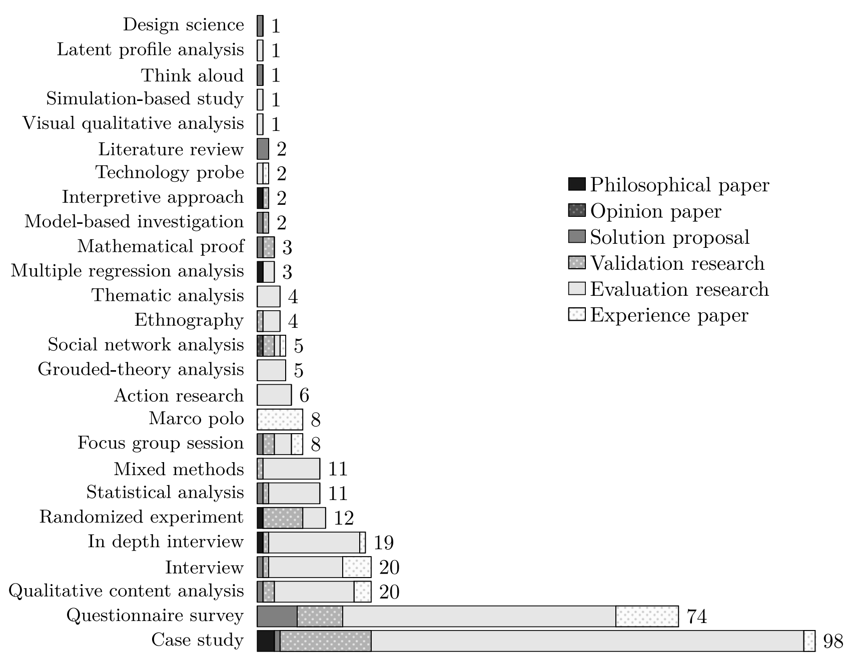

Fig. 7. Research methods and the type of the research.

problem of building a team's integrity rather than investigating the best way to divide a group of students into teams. Group development (3 papers) is partly overlapping in meaning with team building, but team integrity is not necessarily as strongly related to the term. These terms were not used uniformly in the literature. The most apparent issue was with the term "team/group composition", which was often used to describe group formation. We took this issue into account and used our above-mentioned definitions to interpret these terms.

The variety of research on different group work areas reflects the diversity of the research area. Not all areas have attracted much attention, but mentionable are at least the papers focusing on group's performance (15 papers), students' participation in the group (10 papers), group awareness (9 papers), team dynamics (9 papers), the various roles in a group (6 papers), team's physical or virtual environment (6 papers), and the dysfunctional attributes of a team (5 papers). Interestingly, some areas of group work have exclusively taken the form of evaluation research. Such are

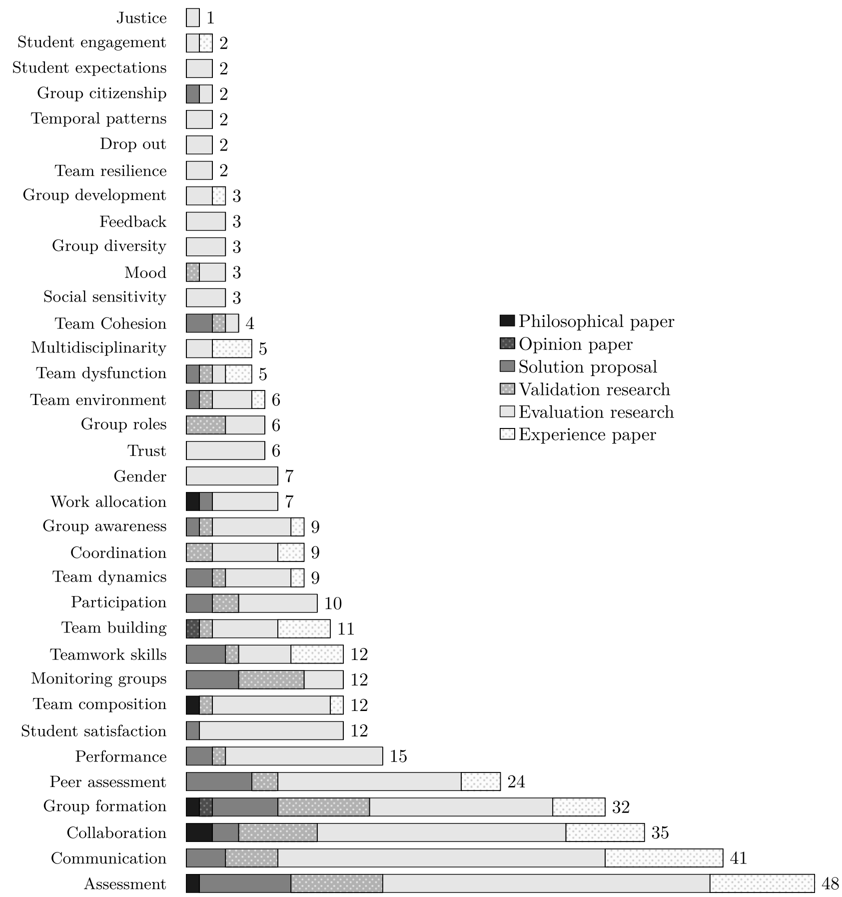

Fig. 8. The areas of group work in selected papers during the last decade. Assessment figures include peer assessment, which is also presented as a separate bar.

e.g., gender-related research (7 papers) and research on trust in a group (6 papers).

Monitoring groups (12 papers) stood out because of its type of research. The papers mostly consisted of solution proposals and validation research instead of evaluation research. Thus, monitoring is mainly studied at the concept level, introducing or testing a novel way of monitoring the activities of a group. Also, there is a set of group work areas, which seem important but have received little attention during the last decade. These include e.g., student *drop out* (2 papers), the related *team resilience* (2 papers), and justice (1 study).

Fig. 9 presents the annual development of the number of papers by the group work areas. Papers focusing on the communication of a group in project education have been published each year for the past decade. The number of publications appears to aggregate more in the early and middle stages of the decade and the interest in the topic seems to have waned slightly over the period 2017–2021. The assessment of group work has attracted researchers in a relatively steady trend, except for a few individual years. Regarding peer assessment, the research interest appears to have grown significantly in 2019. Over the past decade, research on collaboration has been conducted with varying numbers of publications. The topic has been most visible in 2011–2012 and again in 2018 and 2021. Group formation has attracted researchers during the whole decade, with the exception of 2014. Due to the small number of publications, the temporal development of the lesser researched areas might not provide as relevant information as the development of the more studied areas.

The distribution of different group work areas with respect to different forums is illustrated in Fig. 10. Among the journals that stood out in terms of the number of publications, *ACM Transactions on Computing Education* (TOCE), *IEEE Transactions on Education* (ToE), and *Teaching in Higher Education* (TiHE) contained 1–2 papers on the most popular group work areas, but no specific focus on any particular topic was visible. *Journal of Systems and Software* appeared to contain a more diverse set of topics, including lesser researched group work areas such as "student satisfaction" and "group awareness". The most popular conferences, FIE, ITiCSE, and SIGCSE covered almost every group work area found in the mapping. The most prominent clusters of group work areas could be found in the FIE conference, where assessment, communication, and collaboration were well represented.

4.6. Theoretical frameworks

Table 5 illustrates the number of theoretical frameworks, such as theories or concepts, used in the papers. Tuckman's group development theory (Tuckman, 1965) seemed to be the most used theoretical basis for research in the area (13 papers). "Social loafing" (7 papers) represents all types of research except opinion and experience papers. The "Method" (12 papers) in the figure refers to situations in which the theoretical framework used in the study is a method developed, tested, or presented in the study (e.g., Marshall and Gamble, 2015; Sokolowski and Oussena, 2017). "Method" was mainly used in solution proposals and validation studies. Kolb's learning style theory (5 papers) was used in experience and evaluation papers. Big Five personality model[8] (5 papers) has almost exclusively been used in evaluation

8 Also known as the five-factor model of personality (FFM) according to Oxford Bibliographies (Soto and Jackson, 2013).

Fig. 9. The annual development of the research by the different areas of group work.

research, which distinguishes it from another personality framework, Myers-Briggs Type Indicator (5 papers), which was mainly used in validation research. Other theoretical frameworks appear to be used by separate individual research groups and do not form a visible tradition as a theoretical basis for the research area. An example of a such framework is Cognitive Collaborative Model (CCM) (4 papers), which is studied by an individual research group, but counted four times because those papers had different research questions and/or data.

Fig. 11 illustrates how different authors used the theoretical frameworks. Authors are represented by circles with id numbers. The lines between the frameworks and the authors represent the number of times the related author used the related framework. Authors who used only one framework only once are clustered around that framework with no line in-between. Most often authors appeared to use only one theoretical framework.

The usage of different frameworks and concepts as a theoretical basis for researching different group work areas is depicted in Fig. 12. Because a particular study may cover more than one area of group work, the presented numbers do not represent the exact number of papers, but the count of papers the framework appeared together with a specific group work area. The sum of the numbers in horizontal lines may therefore exceed the total number of papers the framework appeared in. Tuckman's theory has been mostly used in group formation research, but it seems to be useful in the research of other group areas too. The Big Five model does not show a similar focus on one area of group work.

5. Discussion

5.1. General observations

Our results enabled us to make conclusions about the foci and nature of the literature on group work research in the context of project-based learning and to identify some research gaps. The majority of the papers were experience or evaluation research. We classified papers as evaluation research with a low threshold if any empirical data analysis was apparent. The most typical study was a case study conducted in a project course that used interviews to gather data. However, in a remarkably small number of papers, systematic qualitative methods (e.g. Grounded theory) or convincing experimental settings (e.g. a randomized comparative experiment) were used.

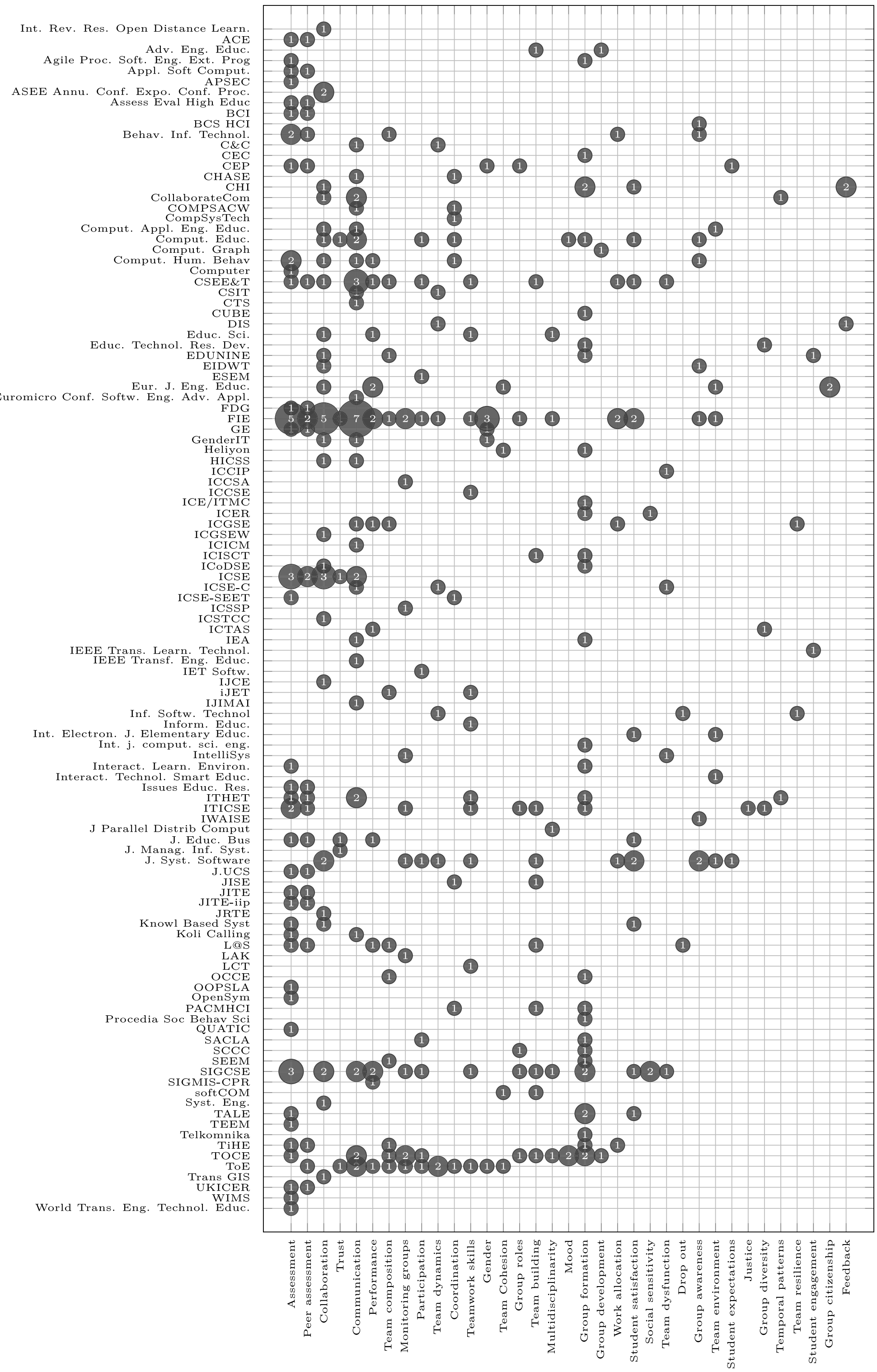

Fig. 10. The representation of the areas of group work in different publication forums.

In Europe, researchers were clearly more focused on graduate-level studies than in North America (see. Table 4). This is probably a natural consequence of the cultural difference between the continents. In North America, undergraduate education in a sense has a stronger position than in Europe, where undergraduate studies are usually a year shorter and students most often pursue graduate studies before starting their careers.

The majority of the papers in the mapping were published at a scientific conference, which may have implications for the nature of the papers in the data. Our observation is, taking into account both included and excluded papers, that a significant portion of the data consisted of experience papers. A typical example of an excluded experience paper was a description of some innovative course module (e.g. Wood et al. (2018)), while the paper by Billings and England (2020) is an example of included experience paper with a sufficient focus on group work. Documenting experiences and observations on a specific course module certainly has its value, but the contribution may remain

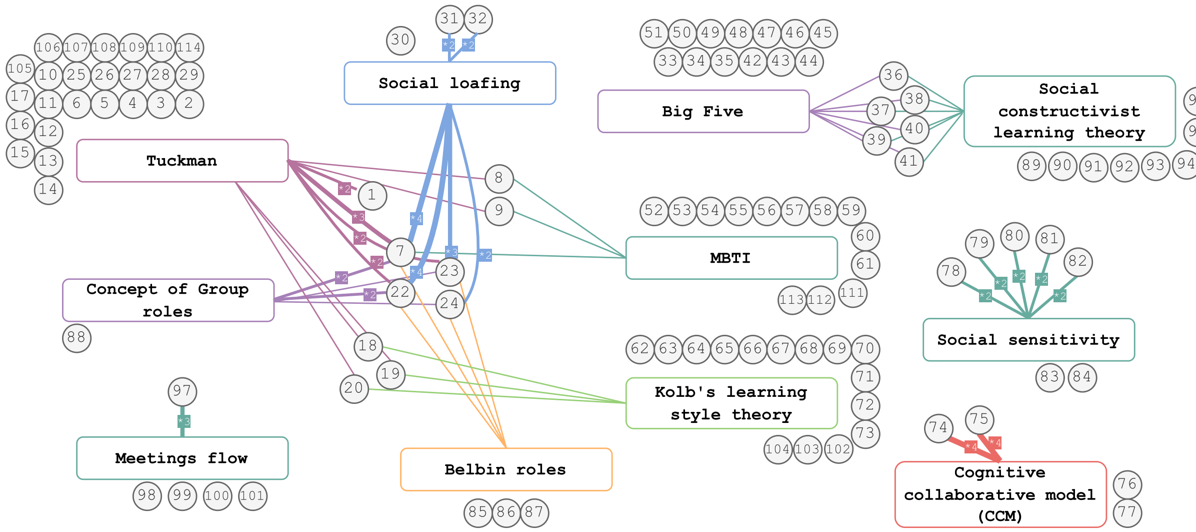

Fig. 11. Usage of theoretical frameworks by different authors (see. author ids in Appendix). Authors who used only one framework are clustered around that framework. Only the frameworks with three or more papers are included. The number in a square means the number of times the particular framework was used by the author.

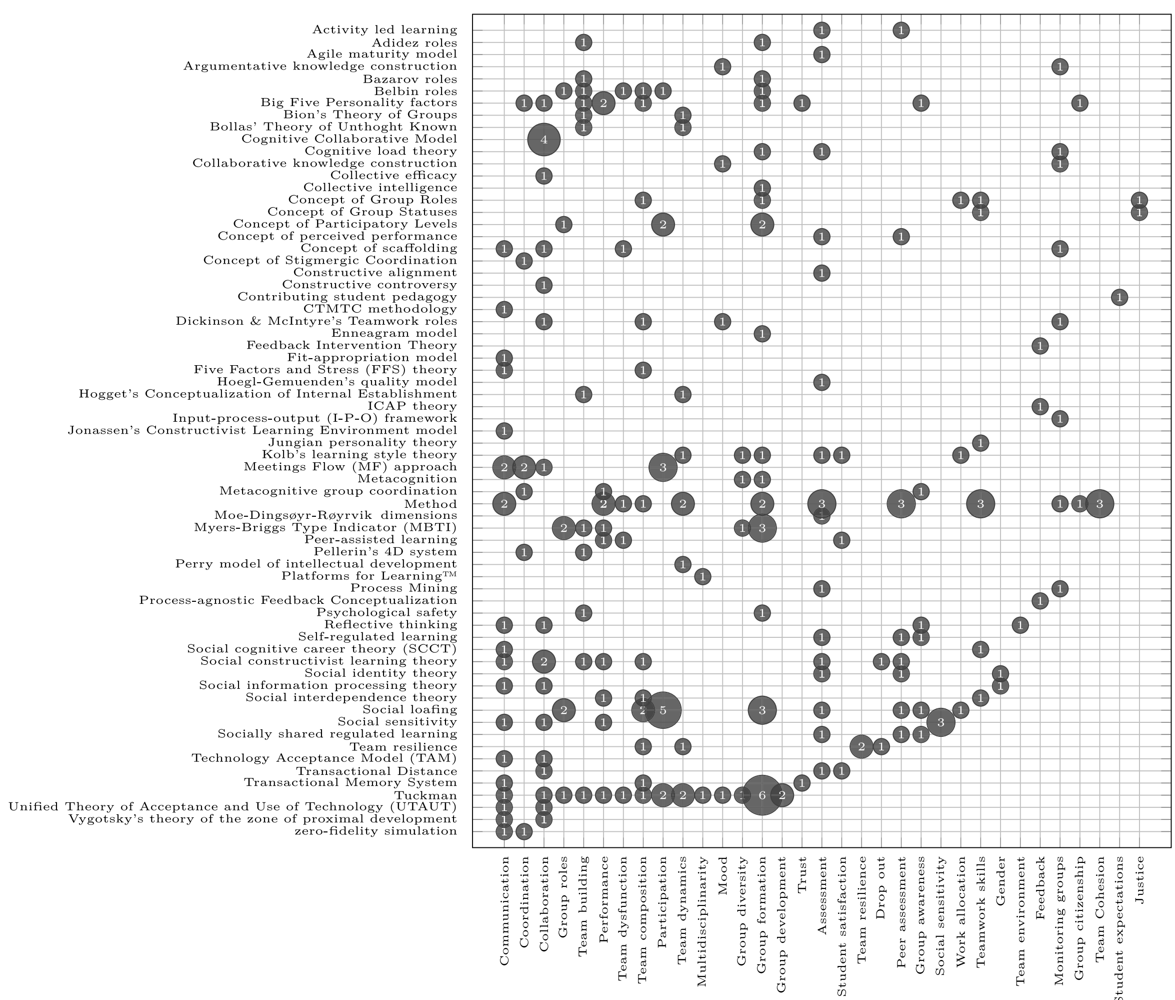

Fig. 12. Theoretical frameworks (vertical axis) per different group work areas (horizontal axis). The bubbles should be interpreted as links between frameworks and group work areas instead of the number of papers because a single paper might be related to multiple frameworks and/or group work areas.

Table 5
Theoretical frameworks and the type of research.

Framework	Phil.	Opin.	Solut.	Valid.	Eval.	Exp.	Total
Tuckman	2	8	1	0	0	2	13
Method	3	2	7	0	0	0	12
Social loafing	1	4	1	1	0	0	7
Myers-Briggs Type Indicator (MBTI)	3	2	0	0	0	0	5
Kolb's learning style theory	0	2	1	0	0	2	5
Big Five Personality factors	0	4	0	0	0	1	5
Cognitive Collaborative Model	2	0	0	2	0	0	4
Social sensitivity	0	3	0	0	0	0	3
Belbin roles	1	0	0	0	1	1	3
Social constructivist learning theory	1	2	0	0	0	0	3
Meetings Flow (MF) approach	1	2	0	0	0	0	3
Team resilience	0	2	0	0	0	0	2
Dickinson & McIntyre's Teamwork roles	1	1	0	0	0	0	2
Concept of Group Roles	0	1	0	1	0	0	2
Process Mining	1	0	1	0	0	0	2
Five Factors and Stress (FFS) theory	0	2	0	0	0	0	2
Vygotsky's theory of the zone of proximal development	0	1	0	0	0	1	2
Concept of scaffolding	0	1	1	0	0	0	2
Reflective thinking	0	2	0	0	0	0	2
Concept of Participatory Levels	0	2	0	0	0	0	2
Social interdependence theory	0	2	0	0	0	0	2
Transactional Distance	0	1	0	0	0	0	1
Fit-appropriation model	0	1	0	0	0	0	1
Activity led learning	0	0	0	0	0	1	1
Peer-assisted learning	0	1	0	0	0	0	1
Argumentative knowledge construction	1	0	0	0	0	0	1
Collaborative knowledge construction	1	0	0	0	0	0	1
Concept of Group Statuses	0	1	0	0	0	0	1
Zero-fidelity simulation	1	0	0	0	0	0	1
Psychological safety	0	1	0	0	0	0	1
Input-process-output (I-P-O) framework	1	0	0	0	0	0	1
Collective intelligence	1	0	0	0	0	0	1
Enneagram model	1	0	0	0	0	0	1
Social information processing theory	0	1	0	0	0	0	1
Piagetian theory of learning	0	0	0	0	0	1	1
Social cognitive career theory (SCCT)	0	0	0	0	0	1	1
Transactional Memory System	0	1	0	0	0	0	1
Adidez roles	0	0	0	0	1	0	1
Bazarov roles	0	0	0	0	1	0	1
Jonassen's Constructivist Learning Environment model	0	1	0	0	0	0	1
Metacognitive group coordination	0	1	0	0	0	0	1
Collective efficacy	0	1	0	0	0	0	1
Social identity theory	0	1	0	0	0	0	1
Constructive controversy	0	1	0	0	0	0	1
Pellerin's 4D system	0	0	0	0	0	1	1
Perry model of intellectual development	0	0	0	0	0	1	1
Socially shared regulated learning	0	1	0	0	0	0	1
Self-regulated learning	0	1	0	0	0	0	1
Metacognition	0	1	0	0	0	0	1
Constructive alignment	1	0	0	0	0	0	1
Jungian personality theory	1	0	0	0	0	0	1
Contributing student pedagogy	0	1	0	0	0	0	1
Cognitive load theory	0	1	0	0	0	0	1
Concept of Stigmergic Coordination	1	0	0	0	0	0	1
Feedback Intervention Theory	0	1	0	0	0	0	1
Process-agnostic Feedback Conceptualization	0	1	0	0	0	0	1
Concept of Retrospectives	0	0	0	0	0	1	1
Platforms for Learning™	0	1	0	0	0	0	1
Bion's Theory of Groups	0	1	0	0	0	0	1
Hogget's Conceptualization of Internal Establishment	0	1	0	0	0	0	1
Bollas' Theory of Unthought Known	0	1	0	0	0	0	1
CTMTC methodology	0	1	0	0	0	0	1
Technology Acceptance Model (TAM)	1	0	0	0	0	0	1
Unified Theory of Acceptance and Use of Technology (UTAUT)	1	0	0	0	0	0	1
Concept of perceived performance	0	0	1	0	0	0	1
Hoegl-Gemuenden's quality model	0	1	0	0	0	0	1
Agile maturity model	0	1	0	0	0	0	1
ICAP theory	0	1	0	0	0	0	1
Moe-Dingsøyr-Røyrvik dimensions	0	1	0	0	0	0	1

thin because of the lack of focus that could be achieved by anchoring the study to a particular theoretical framework.

Bigger conferences, which publish hundreds of papers yearly, appear to have developed diversity in group work domains covered (see Fig. 10). This applies to FIE and SIGCSE in particular, but also ITiCSE when compared with smaller conferences such as Koli Calling. Additionally, it is interesting to speculate that SIGCSE conference being a core CS education conference might have a bit different emphasis than FIE which has 'engineering' in its name. FIE had the biggest volume together with the topic diversity. It is commonly known that for instance research on introductory programming receives lots of attention in SIGCSE.

In light of the results, the most researched topic is group work assessment, half of which concerns peer assessment. Other dominantly visible themes are group formation and team building. These are all topics that the teacher must necessarily take a position on when planning and directing a project course, which we believe, explains their dominance. From this perspective, it is unsurprising that studies on these topics have been published steadily throughout the last decade (see. 9). A second most prominent topic, communication, occurs often together with another group work area (in 65.85% of cases), which may be explained by the observation that communication is quite a general concept and related to many other aspects of teamwork (e.g., most clearly to collaboration). Its popularity is therefore not unexpected. Significantly less attention has been paid to other areas of group work, which may be indicative of how narrowly group work has been studied in CS/SE project education literature.

The top end of the list of theoretical frameworks (see. Table 5) expresses a certain degree of stagnation in the area. Tuckman's group development theory dominates and the next most used theories are far behind, which is in line with the literature review by Riebe et al. (2016), in which the theory was evidently visible in comparison to other group theories. This raises the question of whether the field is moving forward in terms of using novel theoretical perspectives or the Tuckman's theory has an unshakable foothold.

Comparing the field of group work research in CS/SE disciplines to the research in other disciplines may reveal interesting differences. In CS/SE disciplines, the literature seems to lack such concepts of organizational behavior that are used in education research in economy disciplines. For instance, social loafing appeared as a research interest only in seven of the selected papers, but in the economics education literature, it is not difficult to find research dedicated to the topic as a phenomenon, as can be seen from a literature review by Riebe et al. (2016, p. 630), in which it was found as a theme in 17 papers out of a total of 57. Also, such fairness-related concepts as distributive, procedural, and interactive justice could be useful in the study of group work phenomena but are currently absent.

Taking into account the large number of papers selected for the mapping, solid theoretical foundations were found only in a small portion of papers. In 55.11% of the included papers, there was no clear theoretical framework at all. Only a few frameworks have gained a foothold in the literature. Tuckman's stages of group development, Big Five personality traits and Myers–Briggs Type Indicator (MBTI) are famous theoretical foundations that prevailed most visibly in the results but were still not numerous. Moreover, the figures on theoretical frameworks presented (see. Table 5) may not be originated from different papers, but one study might use several of the presented frameworks. Although the personality types appear to be relatively widely used in the SE research, as shown by Barroso et al. (2017), our results (see. Table 5) imply that in educational research they do not seem to populate that much literature. Thus, the theoretical basis of the research area appears somewhat thin over the past decade.

The selection criteria we used ruled out a lot of interesting and topically related papers. The papers that involved group work in the project education context but were not aimed at computing or software engineering audiences did not fit into the mapping. As a result, many laudable papers were excluded, such as the O'Neill et al. (2020) study considering team dynamics.

The work of this systematic mapping study was assisted with a web-based software tool developed for this study, which aimed to reduce the large amount of repetitive and error-prone work. The experiences of collecting the data from various sources to our database, classifying, and keywording the data within the tool were encouraging. The tool was able to reduce the cognitive load and mirror the changes made to the data with a single command to the charts and tables. The developed tool will be reported in a separate publication with its source code.

5.2. Implications for teaching

The mapping study intending to reveal research gaps and direct research may not resonate among teachers as much as with researchers. However, implications for teachings can be commented on. The mappings can serve as educative taxonomy. Firstly, teachers can directly read the taxonomy of group work areas existing in the literature or obtain an instant overview of the theoretical frameworks used. The authors believe that the mappings in this sense help project course teachers to make sense of their teaching settings beyond a particular case and provide a resource for faculties inducting new project course teachers.

Furthermore, the results *separate* areas of group work in the literature: Fig. 8 together with the included mapping information of included papers (see URL in Section 4) provide a resource for teachers with a particular interest in mind (e.g., assessment, group formation, or collaboration). To facilitate and demonstrate readers' access to the results in this manner, selected papers with selected viewpoints, which drew our attention in the sense of lessons learned for teaching during the mapping process, are provided in Table 6.

Lastly, many SE teachers supervise student thesis at several levels of seniority (bachelor, master, and PhD). The results provide information on unused, underused, and used theoretical frameworks, which arguably helps supervisors propose research topics for students; therefore, what is advanced in Section 6 as future research directions applies to teaching as well.

5.3. Limitations

Several limitations are acknowledged. The quality criteria used excluded a part of the "lighter" research from the mapping. Therefore, it is not possible to see directly from the mapping how much of the literature is "light" in nature. On the other hand, the answers to the mapping questions defined in Section 3.2.1 may be more useful because they are not drowned in the large mass of course descriptions and other "light" research. It is also probable that we did not manage to include all the published literature on the topics addressed. Still, we believe that our inspection of 7515 papers offers a useful overview of the research area.

Yet another limitation is that the papers before 2010 were excluded. The reason is that with the present interest in the big picture in place of a narrower study, the number of papers to be screened would have grown beyond what was considered manually possible. This is likely the argument that explains why one can observe systematic literature studies limited to decade-long periods. A speculative argument that supports the present time selection is that computing education research can be said to be increasingly interested in theory-informed research, along with the analyses of how theory has been used in the field (e.g. Malmi et al., 2019). From this perspective, the past decade appears a reasonable focus, and one might see even more course description papers without a dedicated theory-based focus on group work in an older corpus. The present results may help in identifying key papers on particular areas of group work, allowing more focused

Table 6
Selected included papers demonstrating the lessons learned during the mapping.

Paper	Area	Lesson learned	Our reflection
Čavrak et al. (2019)	Team resilience, Drop Out, Team dynamics	Stress from changing requirements can cause low-cohesion distributed teams to drift into a mode in which attention is only given to completing functionalities and educationally valuable processes in a team are lost.	The paper provokes understanding across project types: in our co-located projects, occasions where student(s) start to work beside the project, stressing the project, have caused the focus on functionalities. Echoing (Čavrak et al., 2019), we can also anecdotally recall teams who came together early on and adapted their processes.
Schneider et al. (2015)	Communication, Mood	The paper shows that mood, negativity vs. positivity in teams, is an influential factor. Schneider et al. (2015) argued that mood is perhaps not taken seriously in SEE and SE community.	Teachers' attention is drawn to emotions. The paper can be used as evidence delivered to students, for instance, to caution against mid-project over-optimism that troubles the completion of the project—a result in Schneider et al. (2015).
Heels and Devlin (2019)	Gender, Group roles	Females can tend to select non-technical roles regardless of being strong programmers. This affects peer evaluations since technical roles are valued most.	Teachers need to be alert to the patterns of these kinds and consider the appropriate form of interventions. Group project experiences can be decisive for employers, for why the topic demands continuous attention.
Hastings et al. (2018)	Team building, Group formation	Intentional team formation may not have the expected results compared to random assignment. Psychological safety in groups emerged organically—none of the two grouping strategies used influenced it compared with the other.	The paper outlines the group thing as a complex matter. In our reading, the paper calls attention to the value of dialog in groups, which seems a worthy note because group formation strategies tend to be given ample attention in the field.
Xiao (2013)	Group awareness, Team environment	Collaborative environment for sharing reflections in a group served in different ways; e.g., it helped decision-making and seeing minority opinions in groups. Seeing others' reflections both contributed to and blocked creativity.	It is known that students are heavily occupied by intensive, authentic, project work and, then, not reflection. This paper draws attention to the measures of how continuous reflection could be prompted and considered useful by students.

literature studies to use these as a starting set for snowballing toward older papers, following the guidelines by Wohlin (2014).

6. Future research directions

The mapping conducted suggests at least the following future research considerations:

- An increase in the amount of systematic or theoretically heavy research would be needed. Table 5 provides a starting point by disclosing frameworks that have only been touched upon. The examples below were developed to illustrate how the underused frameworks can inspire more research.
  - The included paper by Wengrowicz et al. (2014) developed a questionnaire for gauging transactional distance (TD) and used it in a setting where students communicated with peers and distance researchers through particular media. TD is historically a theory of distance education (Moore, 2018): It focuses on psychological distance through the factors of how structured the setting is and how much dialog the setting affords. High TD arises from a structured setting with little dialog. The study by Wengrowicz et al. (2014), and the use of TD, could be extended to research on group dynamics and performance. In particular, it is suggested that TD could be used to understand how students adopt a software process and how that affects their performance. For instance, if work is simply divided among group members (a highly structured setting) with little correspondence until the group is forced to integrate individual efforts for a shared outcome (a low-dialog setting), TD would be high and, following from the theory, performance would depend on how prepared for autonomy group members were. Researchers could analyze, for instance, how students manage to adopt a highly collaborative Scrum and how they perform in this respect. Additionally, TD appears a promising framework because communication was a clearly visible domain of group work in the present mapping.
  - The included paper by Lin (2018) studied the effects of group awareness and peer evaluation on socially shared regulated learning (SSRL; see Hadwin et al. (2017)) with the promising results of increased attention to collaboration and reduced free riding. The study was conducted in the context of CSCL in which team interactions can be tracked and used as information for group awareness. However, it is suggested that SSRL could inform several kinds of group intervention attempts regardless of whether the setting is either online or face2face. For instance, it may be an appropriate term to explain the increased actions within SE student teams when – in another included study – a justice-based dialogic intervention caused teams to collaborate for concrete corrective actions (Isomöttönen and Ritvos, 2021).
  - Justice is arguably implicitly present in many papers (e.g., those on assessment and social loafing), but justice as an explicit theoretical framework appears to be largely missing at the level of group pedagogy. One article was identified with justice as its main focus (Fig. 8) but without the adoption of a particular theory. Justice has recently received attention in computing education special issues, while these important studies focus on topics of justice at a more structural level (see, Ryoo et al., 2022). The present authors have attempted (work-in-progress paper) group analysis by adopting Adams' Theory of Equity as a tool to understand students' experiences of group work (Kokkoniemi and Isomöttönen, 2020). Adam's theory is mostly related

to distributive justice: how reward is divided in light of efforts. The mapping suggests there is room for analyzing various types of justice (e.g., distributive, procedural, interactional) and, it is suggested here, how these constructs explain team satisfaction and group member behaviors.

– An identified but somewhat unnoticed framework in the included papers appears to be Belbin Team Roles. Of the included papers that were judged to have Belbin roles as part of their framework (Marshall et al., 2016; Leung, 2017; Zvereva and Muravyov, 2019), the first two can be said to include attention to Belbin theory. Marshall et al. used Belbin roles to introduce students to a pedagogically intended, challenging group experience (Belbin broken vs. Belbin balanced teams), whereas research-wise more attention was given to the analysis of students' participatory levels. On reflection, it could be asked if the participation levels, or justice-related behaviors in group work in general, are more crucial determinants of students' group experiences than a particular role scheme. Hence, more clarity about the Belbin roles in SE student groups is welcome. Leung (2017) informed students of Belbin roles and used them as a framework for reflection at the end of the projects. However, more evaluative research is needed on how useful students see such education. These evaluations should analyze Belbin scheme with different student team sizes.
– Cook et al. (2019) used ICAP framework for understanding the nature of peer feedback processes in a project-based setting. The ICAP framework defines different cognitive engagements for a learner: interactive, constructive, active, and passive (Chi and Wylie, 2014). The use of the framework could be extended to inquiries into individual students' learning during realistic group projects where autonomy is expected. For instance, varying skill levels in a programming group may dictate a passive, kind of follower role for inexperienced students (Isomöttönen and Tirronen, 2013). The proposed research could be conducted at various stages of a curriculum to gain a longitudinal understanding of how an individual's learning relates to their group. Evaluations based on the ICAP framework would entail comparability between such studies.
– Escudeiro et al. (2011) used Social Cognitive Career Theory (SCCT) for interpreting their course evaluation. This theory is based on Bandura's social cognitive theory and links the interplay among interest, choice, and performance with learning experiences (Lent et al., 2002, pp. 264–267). The students' learning experiences are affected by how they perceive their performance and can influence their interests and what activities they choose to invest in Lent et al. (2002, p. 267). In the present context, it is suggested that this theory persuades longitudinal studies on the effects of students' group work experiences: We discussed students' negative group experiences affecting their attitudes and, relatedly, the students' fear of unpredictability in Section 2.3. It is these kinds of cross-curricular effects that computing educators should look at in more detail with help of SCCT. Additionally, an interesting question is how particular group experiences (e.g., justice sentiments) mold students' prospects and potentially influence their choices even in subsequent job settings.

- The few apparently prominent frameworks (Tuckman, social loafing, MBTI, Big Five, and Kolb) were relatively little used when compared with the number of papers included. This observation also calls for more theoretically-based research and suggests that theorizations from other disciplines should be integrated into SE student projects research more richly.
- The present mapping study paves a way for detailed review studies. A follow-up review should analyze *how* theories or conceptual frameworks have been used (related to both Tables 5 and 11). A preliminary inspection shows that this varies. The proposed review should analyze all included papers although a study would not raise a particular theory or conceptual framework. In such cases, it is worth analyzing if the related work section of the paper informed the study setting or was reflected on regarding reported results or experiences. Reviews synthesizing the educational benefits and challenges of the use of a particular theoretical framework are also available.
- A high number of evaluation papers with case study characteristics were identified. An interesting question is how systematic and data-rich research can be identified in these papers; how many of these papers represent a small step forward from experience reporting, e.g., by reviewing some student course feedback data? Developing such synthesis would further illuminate the nature of the field while helping to assess the used research type categorization for the computing education area.
- A remark made during the selection process was that the papers were often aimed at a wide software engineering audience, which was likely to affect the pedagogical depth of the presentation. A contribution to SEE research might become available with a clearer educational focus. To this end, one option could be a double publication, in which a separate version of the study would be published with a theoretically grounded pedagogical point of view.

In addition to above proposals, the readers can take advantage of the results (Fig. 8 and Table 5) by checking if their emerging research interests have been addressed.

7. Conclusion

Out of 7515 initial search results generated by the search strategy, 225 eligible papers were selected. Among the areas of group work, communication, assessment and group formation have gained the most attention; most of the papers were experience or evaluation research; The papers were most often published in a scientific conference; For a considerable number of papers, the nature of the research can be regarded as not that serious or theoretically grounded; Theoretical frameworks were used relatively little with respect to the number of papers published; The research appears to be quite narrowly focused on a few areas. Our observations of the selected literature warrant the following conclusions: Firstly, we conclude a need for more focused CS/SE project education research, in which project education is inspected from particular perspectives. This would create identifiable lines of research and structure the research area. Secondly, those theoretical frameworks that have currently been only touched upon should receive more attention. Thirdly, the research area could be precipitated by novel theoretical perspectives. For practitioners, those implementing a group project course can benefit from a large amount of literature on assessment and group formation, which are issues on which the teacher must take a position. The lessons learned examples attempt to pinpoint essential areas of group work for teachers. Generally, the present results outlining the field in a structured way can facilitate research-based teaching and thesis supervision in different areas of project education.

Table A.1

ID	Name	ID	Name	ID	Name
1	David L. Largent	39	Michael Smithson	77	Raghvinder S. Sangwan
2	Kurt Schneider	40	Lynette Johns-Boast	78	Lisa Bender
3	Olga Liskin	41	Dirk van Rooy	79	Gursimran Walia
4	Hilko Paulsen	42	Marguerite Doman	80	Krishna Kambhampaty
5	Simone Kauffeld	43	Andrew Besmer	81	Kendall E. Nygard
6	Chris Lüer	44	Anne Olsen	82	Travis E. Nygard
7	Vreda Pieterse	45	Amir Mujkanovic	83	Lisa L. Lacher
8	Mpho Leeu	46	Andreas Bollin	84	Cydnee Biehl
9	Marko van Eekelen	47	David Bell	85	Wai Sze Leung
10	Maria-Iuliana Dascalua	48	Tracy Hall	86	Olga Zvereva
11	Ana-Maria Dumitrachea	49	Jo Erskine Hannay	87	Artyom Muravyov
12	Melania Comanb	50	Dietmar Pfahl	88	Ville Isomöttönen
13	Alin Moldoveanu	51	Silvia Teresita Acuna	89	Thomas Staubitz
14	Philipp Neumanna	52	Luis G. Martínez	90	Christoph Meinel
15	Christoph Kowitza	53	Guillermo Licea	91	Fatma Cemile Serçe
16	Felix Schrannerb	54	Antonio Rodríguez-Díaz	92	Kathleen Swigger
17	Dmitrii Azarnykhb	55	Juan R. Castro	93	Ferda Nur Alpaslan
18	Mark Fontenot	56	Lenuta Alboaie	94	Robert Brazile
19	Katherine Canales	57	Mircea-Florin Vaida	95	George Dafoulas
20	Andrew N. Quicksall	58	Diana Pojar	96	Victor Lopeze
21	*(unassigned ID)*	59	Luis Silvestre	97	Chung-Yang Chen
22	Lisa Thompson	60	Sergio F. Ochoa	98	Ya-Chun Hong
23	Linda Marshall	61	Maira Marques	99	Pei-Chi Chen
24	Dina M. Venter	62	Yekaterina Kharitonova	100	Kao-Chiuan Teng
25	Marco Kuhrmann	63	Yi Luo	101	P. Pete Chong
26	Jürgen Münch	64	Jeho Park	102	Yi Meng Lau
27	Kearney, Kerri S.	65	Jian Chen	103	Kyong Jin Shim
28	Damron, Rebecca	66	Guoyong Qiu	104	Swapna Gottipati
29	Sohoni, Sohum	67	Liu Yuan	105	Nickolas Falkner
30	Jian-Wei Lin	68	Li Zhang	106	Otto Seppälä
31	Ilenia Fronza	69	Gang Lu	107	Tapio Auvinen
32	Xiaofeng Wang	70	Maria Kyprianidou	108	Ville Karavirta
33	Shuddha Chowdhury	71	Stavros Demetriadis	109	Petri Ihantola
34	Charles Walter	72	Thrasyvoulos Tsiatsos	110	Arto Hellas
35	Rose Gamble	73	Andreas Pombortsis	111	Budi Laksono Putro
36	Conal Monaghan	74	Joanna F. DeFranco	112	Yusep Rosmansyah
37	Boris Bizumic	75	Colin J. Neill	113	Suhardi
38	Katherine Reynolds	76	Roy B. Clariana	114	Yasar Güneri Sahin

CRediT authorship contribution statement

Mikko Kokkoniemi: Conceptualization, Methodology, Software, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Ville Isomöttönen:** Conceptualization, Methodology, Investigation, Writing – original draft, 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

We share a link to our data in the beginning of the "results" section.

Appendix. Author IDs for Fig. 11

See Table A.1.

References

Adderley, K., et al., 1975. Project Methods in Higher Education. In: Research into higher education monographs, vol. 24, Society for research into higher education.

Ahmad, M.O., Liukkunen, K., Markkula, J., 2014. Student perceptions and attitudes towards the software factory as a learning environment. In: IEEE Global Engineering Education Conference. EDUCON, IEEE, pp. 422–428. http://dx.doi.org/10.1109/EDUCON.2014.6826129.

Allan, E.G., 2016. I hate group work!: Addressing students' concerns about small-group learning. InSight: J. Sch. Teach. 11, 81–89, https://eric.ed.gov/?id=EJ1110139.

Apiola, M., Lattu, M., Pasanen, T.A., 2012. Creativity-supporting learning environment—CSLE. Trans. Comput. Educ. 12, 11:1–11:25. http://dx.doi.org/10.1145/2275597.2275600.

Barr, M., Parkinson, J., 2019. Developing a work-based software engineering degree in collaboration with industry. In: Proceedings of the 2019 Conference on United Kingdom & Ireland Computing Education Research. pp. 1–7.

Barroso, A.S., Madureira, S.J., Soares, M.S., do Nascimento, R.P.C., 2017. Influence of human personality in software engineering - a systematic literature review. In: Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 1: ICEIS. INSTICC. SciTePress, pp. 53–62. http://dx.doi.org/10.5220/0006292000530062.

Bastarrica, M.C., Perovich, D., Gutierrez, F.J., Marques, M., 2019. A grading schema for reinforcing teamwork quality in a capstone course. In: IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion). IEEE/ACM, pp. 276–277. http://dx.doi.org/10.1109/ICSE-Companion.2019.00112.

Billings, S., England, M., 2020. First year computer science projects at coventry university: Activity-led integrative team projects with continuous assessment. In: Proceedings of the 4th Conference on Computing Education Practice 2020. Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3372356.3372358.

Borges, S., Mizoguchi, R., Bittencourt, I.I., Isotani, S., 2018. Group formation in cscl: A review of the state of the art. In: Cristea, A.I., Bittencourt, I.I., Lima, F. (Eds.), Higher Education for All. from Challenges To Novel Technology-Enhanced Solutions. Springer International Publishing, Cham, pp. 71–88. http://dx.doi.org/10.1007/978-3-319-97934-2_5.

Bothe, K., 2001. Reverse engineering: The challenge of large-scale real-world educational projects. In: Software Engineering Education and Training. Proceedings. 14th Conference on. IEEE Computer Society, Los Alamitos, CA, pp. 115–126. http://dx.doi.org/10.1109/CSEE.2001.913828.

Burdett, J., 2007. Degrees of separation—balancing intervention and independence in group work assignments. Aust. Educ. Res. 34, 55–71. http://dx.doi.org/10.1007/BF03216850.

Burge, J., Gannod, G., 2009. Dimensions for categorizing capstone projects. In: Software Engineering Education and Training. Proceedings. 22nd Conference on. IEEE Computer Society, Los Alamitos, CA, pp. 166–173. http://dx.doi.org/10.1109/CSEET.2009.37.

Čavrak, I., Bosnić, I., Ciccozzi, F., Mirandola, R., 2019. Resilience of distributed student teams to stress factors: A longitudinal case-study. Inf. Softw. Technol. 114, 258–274. http://dx.doi.org/10.1016/j.infsof.2019.05.011.

Chi, M.T., Wylie, R., 2014. The icap framework: Linking cognitive engagement to active learning outcomes. Educ. Psychol. 49, 219–243. http://dx.doi.org/10.1080/00461520.2014.965823.

Christensen, K., Rundus, D., 2003. The capstone senior design course: An initiative in partnering with industry. In: ASEE/IEEE Frontiers in Education, 33rd Annual, Vol. 3. IEEE, pp. S2B – 12–17. http://dx.doi.org/10.1109/FIE.2003.1265941.

Clark, N., Davies, P., Skeers, R., 2005. Self and peer assessment in software engineering projects. In: Proceedings of the 7th Australasian Conference on Computing Education - Volume 42. Australian Computer Society, Darlinghurst, Australia, pp. 91–100, https://dl.acm.org/doi/abs/10.5555/1082424.1082436.

Clear, T., 2002. A diagnostic technique for addressing group performance in capstone projects. SIGCSE Bull. 34, 196. http://dx.doi.org/10.1145/637610.544475.

Clear, T., Beecham, S., Barr, J., Daniels, M., McDermott, R., Oudshoorn, M., Savickaite, A., Noll, J., 2015. Challenges and recommendations for the design and conduct of global software engineering courses: A systematic review. In: Proceedings of the 2015 ITiCSE on Working Group Reports. ACM, pp. 1–39. http://dx.doi.org/10.1145/2858796.2858797.

Clear, T., Goldweber, M., Young, F.H., Leidig, P.M., Scott, K., 2001. Resources for instructors of capstone courses in computing. In: ITiCSE-WGR '01: Working Group Reports from ITiCSE on Innovation and Technology in Computer Science Education. ACM, New York, NY, pp. 93–113. http://dx.doi.org/10.1145/572139.572179.

Cook, A., Hammer, J., Elsayed-Ali, S., Dow, S., 2019. How guiding questions facilitate feedback exchange in project-based learning. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, pp. 1–12.

Costa, A., Ramos, F., Perkusich, M., Dantas, E., Dilorenzo, E., Chagas, F., Meireles, A., Albuquerque, D., Silva, L., Almeida, H., Perkusich, A., 2020. Team formation in software engineering: A systematic mapping study. IEEE Access 8, 145687–145712. http://dx.doi.org/10.1109/ACCESS.2020.3015017.

Cruz, W.M., Isotani, S., 2014. Group formation algorithms in collaborative learning contexts: A systematic mapping of the literature. In: Baloian, N., Burstein, F., Ogata, H., Santoro, F., Zurita, G. (Eds.), Collaboration and Technology. Springer International Publishing, Cham, pp. 199–214. http://dx.doi.org/10.1007/978-3-319-10166-8_18.

Daniels, M., 2011. Developing and Assessing Professional Competencies: A Pipe Dream? (Ph.D. thesis). Uppsala University, Uppsala.

Daniels, M., Asplund, L., 1999. Full scale industrial project work, a one semester course. In: ASEE/IEEE Frontiers in Education Conference, 29th Annual. IEEE, pp. 11B2/7–11B2/9.

DeFranco, J.F., Laplante, P.A., 2017. Review and analysis of software development team communication research. IEEE Trans. Prof. Commun. 60, 165–182. http://dx.doi.org/10.1109/TPC.2017.2656626.

Dugan, J., 2011. A survey of computer science capstone course literature. Comput. Sci. Educ. 21, 201–267. http://dx.doi.org/10.1080/08993408.2011.606118.

Dybå, T., Dingsøyr, T., Hanssen, G.K., 2007. Applying systematic reviews to diverse study types: An experience report. In: ESEM. ACM / IEEE Computer Society, pp. 225–234. http://dx.doi.org/10.1109/ESEM.2007.59.

Escudeiro, N., Escudeiro, P., Barata, A., Lobo, C., 2011. Enhancing students team work and communication skills in international settings. In: International Conference on Information Technology Based Higher Education and Training. IEEE, pp. 1–8. http://dx.doi.org/10.1109/ITHET.2011.6018683.

Falkner, K., Falkner, N.J., Vivian, R., 2013. Collaborative learning and anxiety: A phenomenographic study of collaborative learning activities. In: Proceeding of the 44th ACM Technical Symposium on Computer Science Education. Association for Computing Machinery, New York, NY, USA, pp. 227–232. http://dx.doi.org/10.1145/2445196.2445268.

Farrell, V., Ravalli, G., Farrell, G., Kindler, P., Hall, D., 2012. Capstone project: Fair, just and accountable assessment. In: Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education. ACM, New York, NY, http://dx.doi.org/10.1145/2325296.2325339.

Fernandez, A., Insfrán, E., Abrahão, S., 2011. Usability evaluation methods for the web: A systematic mapping study. Inf. Softw. Technol. 53, 789–817. http://dx.doi.org/10.1016/j.infsof.2011.02.007.

Figl, K., 2010. A systematic review of developing team competencies in information systems education. J. Inf. Syst. Educ. 21, 323–337.

Fincher, S., Petre, M., Clark, M. (Eds.), 2001. Computer Science Project Work: Principles and Pragmatics. Springer-Verlag, London.

Fronza, I., Wang, X., 2017. Towards an approach to prevent social loafing in software development teams. In: ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM, IEEE, pp. 241–246. http://dx.doi.org/10.1109/ESEM.2017.37.

Hadwin, A., Järvelä, S., Miller, M., 2017. Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In: Handbook of Self-Regulation of Learning and Performance. Routledge, New York, NY, pp. 83–106.

Harris, A.M., Gómez-Zará, D., DeChurch, L.A., Contractor, N.S., 2019. Joining together online: The trajectory of cscw scholarship on group formation. In: Proc. ACM Hum.-Comput. Interact., Vol. 3. http://dx.doi.org/10.1145/3359250.

Hastings, E.M., Jahanbakhsh, F., Karahalios, K., Marinov, D., Bailey, B.P., 2018. Structure or nurture? Proc. ACM Hum.-Comput. Interact. 2, 1–21. http://dx.doi.org/10.1145/3274337.

Hayes, J., Lethbridge, T., Port, D., 2003. Evaluating individual contribution toward group software engineering projects. In: Software Engineering, 2003. Proceedings. 25th International Conference on. pp. 622–627. http://dx.doi.org/10.1109/ICSE.2003.1201246.

Heels, L., Devlin, M., 2019. Investigating the role choice of female students in a software engineering team project. In: Proceedings of the 3rd Conference on Computing Education Practice. ACM, New York, NY, pp. 2:1–2:4. http://dx.doi.org/10.1145/3294016.3294028.

Heikkinen, J., Isomöttönen, V., 2015. Learning mechanisms in multidisciplinary teamwork with real customers and open-ended problems. Eur. J. Eng. Educ. 40, 653–670. http://dx.doi.org/10.1080/03043797.2014.1001818.

Helle, L., Tynjälä, E., 2006. Project-based learning in post-secondary education — theory, practice and rubber sling shots. High. Educ. 51, 287–314. http://dx.doi.org/10.1007/s10734-004-6386-5.

Henttonen, K., 2010. Exploring social networks on the team level—a review of the empirical literature. J. Eng. Technol. Manage. 27, 74–109. http://dx.doi.org/10.1016/j.jengtecman.2010.03.005.

Hernández-March, J., Martín del Peso, M., Leguey, S., 2009. Graduates' skills and higher education: The employers' perspective. Tert. Educ. Manage. 15, 1–16. http://dx.doi.org/10.1080/13583880802699978.

Hinds, P.J., Carley, K.M., Krackhardt, D., Wholey, D., 2000. Choosing work group members: Balancing similarity, competence, and familiarity. Organ. Behav. Hum. Decis. Process. 81, 226–251. http://dx.doi.org/10.1006/obhd.1999.2875.

Iacob, C., Faily, S., 2019. Exploring the gap between the student expectations and the reality of reamwork in undergraduate software engineering group projects. J. Syst. Softw. 157, 110393. http://dx.doi.org/10.1016/j.jss.2019.110393.

Isomöttönen, V., 2011. Theorizing a one-semester real customer student software project course. In: Jyväskylä Studies in Computing, Vol. 140 (Ph.D. thesis). University of JyväSkylä, http://urn.fi/urn:isbn:978-951-39-4534-3.

Isomöttönen, V., 2014. Making group processes explicit to student: A case of justice. In: Proceedings of the 2014 Conference on Innovation and Technology in Computer Science Education. ACM, pp. 195–200. http://dx.doi.org/10.1145/2591708.2591717.

Isomöttönen, V., Daniels, M., Cajander, Å., Pears, A., McDermott, R., 2019. Searching for global employability ACM Trans. Comput. Educ. – Spec. Issue Glob. Softw. Eng. Educ. 19, 11:1–11:29. http://dx.doi.org/10.1145/3277568.

Isomöttönen, V., Ritvos, E., 2021. Digging into group establishment: Intervention design and evaluation. J. Syst. Softw. vol. 178, 110974. http://dx.doi.org/10.1016/j.jss.2021.110974.

Isomöttönen, V., Ryynänen, S., Mononen, N., 2018. Method matters: Reflections from student-made mapping studies. In: 2018 IEEE Frontiers in Education Conference. FIE, pp. 1–9. http://dx.doi.org/10.1109/FIE.2018.8659240.

Isomöttönen, V., Tirronen, V., 2013. Teaching programming by emphasizing self-direction: How did students react to active role required of them? Trans. Comput. Educ. vol. 13, 6:1–6:21.

Jahanbakhsh, F., Fu, W.T., Karahalios, K., Marinov, D., Bailey, B.P., 2017. You want me to work with who?: Stakeholder perceptions of automated team formation in project-based courses. In: Mark, G., Fussell, S.R., Lampe, C., schraefel, m.c., Hourcade, J.P., Appert, C., Wigdor, D. (Eds.), CHI. ACM, pp. 3201–3212. http://dx.doi.org/10.1145/3025453.3026011.

Johansson, C., Molin, P., 1996. Maturity, motivation and effective learning in projects—benefits from using industrial clients. In: SEHE '95, Proceedings of the Second International Conference on Software Engineering in Higher Education II. In: Computational Mechanics, Billerica, MA, pp. 99–106.

Jun, H., 2010. Improving undergraduates' teamwork skills by adapting project-based learning methodology. In: 5th International Conference on Computer Science & Education. IEEE. IEEE, pp. 652–655. http://dx.doi.org/10.1109/ICCSE.2010.5593527.

Kamau, C., Spong, A., 2015. A student teamwork induction protocol. Stud. High. Educ. vol. 40, 1273–1290. http://dx.doi.org/10.1080/03075079.2013.879468.

Khoo, E., Zegwaard, K., Adam, A., 2020. Employer and academic staff perceptions of science and engineering graduate competencies. Australas. J. Eng. Educ. vol. 25, 103–118.

Kilpatrick, W., 1918. The project method. Teach. Coll. Rec. 19, 319–335.

Kitchenham, B.A., Budgen, D., Brereton, O.P., 2011. Using mapping studies as the basis for further research–a participant-observer case study. Inf. Softw. Technol. 53, 638–651. http://dx.doi.org/10.1016/j.infsof.2010.12.011.

Kitchenham, B., Charters, S., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE, 2007-001.

Knoke, P.J., 1991. Medium size project model: Variations on a theme. In: Proceedings of the SEI Conference on Software Engineering Education. Springer-Verlag, London, pp. 5–24.

Knoll, M., 2012. I had made a mistake: William h, kilpatrick and the project method. Teach. Coll. Rec. 114, 1–45.

Kokkoniemi, M., Isomöttönen, V., 2020. Project education and adams' theory of equity. In: 2020 IEEE Frontiers in Education Conference. FIE, pp. 1–5. http://dx.doi.org/10.1109/FIE44824.2020.9274126.

Kokkoniemi, M., Isomöttönen, V., 2023. Dataset and summary of analysis for "Systematic Mapping Study on Group Work Research in Computing Education Projects" http://dx.doi.org/10.17632/dcpymshp8k.1.

Latane, B., Williams, K., Harkins, S., 1979. Many hands make light the work: The causes and consequences of social loafing. J. Personal. Soc. Psychol. vol. 37, 822. http://dx.doi.org/10.1037/0022-3514.37.6.822.

Lent, R.W., Brown, S.D., Hackett, G., et al., 2002. Social Cognitive Career Theory, Vol. 4. pp. 255–311.

Leung, W.S., 2017. Bad blood: managing toxic relationships through belbin roles for first year software engineering students. In: Proceedings of the 3rd International Conference on Communication and Information Processing. ACM, New York, NY, pp. 82–86. http://dx.doi.org/10.1145/3162957.3163010.

Lin, J.W., 2018. Effects of an online team project-based learning environment with group awareness and peer evaluation on socially shared regulation of learning and self-regulated learning. Behav. Inf. Technol. 37, 445–461. http://dx.doi.org/10.1080/0144929X.2018.1451558.

Livingstone, D., Lynch, K., 2000. Group project work and student-centred active learning: Two different experiences. Stud. High. Educ. 25, 325–345. http://dx.doi.org/10.1080/713696161.

Majanoja, A.M., Vasankari, T., 2018. Reflections on teaching software engineering capstone course. In: McLaren, B.M., Reilly, R., Zvacek, S., Uhomoibhi, J.O. (Eds.), CSEDU (2). SciTePress, pp. 68–77. http://dx.doi.org/10.5220/0006665600680077.

Malmi, L., Sheard, J., Kinnunen, P., Sinclair, J., 2019. Computing education theories: What are they and how are they used? In: Proceedings of the ACM Conference on International Computing Education Research. ACM, New York, NY, pp. 187–197. http://dx.doi.org/10.1145/3291279.3339409.

Marques, M., Ochoa, S.F., 2014. Improving teamwork in students software projects. In: IEEE 27th Conference on Software Engineering Education and Training (CSEE T). pp. 99–108. http://dx.doi.org/10.1109/CSEET.2014.6816787.

Marshall, A., Gamble, R.F., 2015. Gauging influence in software development teams. In: FIE. IEEE Computer Society, pp. 1–8. http://dx.doi.org/10.1109/FIE.2015.7344106.

Marshall, L., Pieterse, V., Thompson, L., Venter, D.M., 2016. Exploration of participation in student software engineering teams. ACM Trans. Comput. Educ. 16, 1–38. http://dx.doi.org/10.1145/2791396.

McEwan, D., Ruissen, G.R., Eys, M.A., Zumbo, B.D., Beauchamp, M.R., 2017. The effectiveness of teamwork training on teamwork behaviors and team performance: A systematic review and meta-analysis of controlled interventions. PLoS One 12, 1–23. http://dx.doi.org/10.1371/journal.pone.0169604.

Mochol, M., Tolksdorf, R., 2009. Praxis-oriented teaching via client-based software projects. In: Pears, A., Schulte, C. (Eds.), 9th Koli Calling: International Conference on Computing Education Research. Uppsala University, Uppsala, Sweden, pp. 31–34.

Moore, M.G., 2018. The Theory of Transactional Distance, fourth ed. Routledge, New York, NY, pp. 32–46.

Morgan, A., 1983. Theoretical aspects of project-based learning in higher education. Br. J. Educ. Technol. 14, 66–78. http://dx.doi.org/10.1111/j.1467-8535.1983.tb00450.x.

Nand, R., Sharma, A., Reddy, K., 2018. Skill-based group allocation of students for project-based learning courses using genetic algorithm: Weighted penalty model. In: TALE. IEEE, pp. 394–400. http://dx.doi.org/10.1109/TALE.2018.8615127.

Oakley, B., Felder, R.M., Brent, R., Elhajj, I., 2004. Turning student groups into effective teams. J. Stud. Cent. Learn. 2, 9–34.

O'Neill, T.A., Pezer, L., Solis, L., Larson, N., Maynard, N., Dolphin, G.R., Brennan, R.W., Li, S., 2020. Team dynamics feedback for post-secondary student learning teams: introducing the "bare care" assessment and report. Assess. Eval. High. Educ. vol. 45, 1121–1135. http://dx.doi.org/10.1080/02602938.2020.1727412.

Pears, A., Seidman, S., Eney, C., Kinnunen, P., Malmi, L., 2005. Constructing a core literature for computing education research. ACM SIGCSE Bull. vol. 37, 152–161. http://dx.doi.org/10.1145/1113847.1113893.

Pérez, B., Rubio, Á.L., 2020. A project-based learning approach for enhancing learning skills and motivation in software engineering. In: Proceedings of the 51st ACM Technical Symposium on Computer Science Education. pp. 309–315. http://dx.doi.org/10.1145/3328778.3366891.

Perrenet, J.C., Bouhuijs, P.A.J., Smits, J.G.M.M., 2000. The suitability of problem-based learning for engineering education: Theory and practice. Teach. High. Educ. 5, 345–358. http://dx.doi.org/10.1080/713699144.

Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M., 2008. Systematic mapping studies in software engineering. In: EASE. pp. 68–77.

Petersen, K., Gencel, C., 2013. Worldviews, research methods, and their relationship to validity in empirical software engineering research. In: 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement. IEEE Computer Society, pp. 81–89. http://dx.doi.org/10.1109/IWSM-Mensura.2013.22.

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.

Pieterse, V., Thompson, L., 2006. A model for successful student teams. In: The 36th Annual Conference of the Southern African Computer Lecturers' Association. SACLA, pp. 195–205.

Pieterse, V., Thompson, L., 2010. Academic alignment to reduce the presence of 'social loafers' and 'diligent isolates' in student teams. Teach. High. Educ. vol. 15, 355–367. http://dx.doi.org/10.1080/13562517.2010.493346.

Poort, I., Jansen, E., Hofman, A., 2022. Does the group matter? effects of trust, cultural diversity, and group formation on engagement in group work in higher education. High. Educ. Res. Dev. 41, 511–526. http://dx.doi.org/10.1080/07294360.2020.1839024.

Reis, R., Isotani, S., Rodriguez, C., Lyra, K., Jaques, P., Bittencourt, I., 2018. Affective states in computer-supported collaborative learning: Studying the past to drive the future. Comput. Educ. 120, 29–50. http://dx.doi.org/10.1016/j.compedu.2018.01.015.

Riebe, L., Girardi, A., Whitsed, C., 2016. A systematic literature review of teamwork pedagogy in higher education. Small Group Res. 47, 619–664. http://dx.doi.org/10.1177/1046496416665221.

Ryoo, J.J., Santo, R., Lachney, M., 2022. Introduction to the special issue on justice-centered computing education, part 2. ACM Trans. Comput. Educ. vol. 22, 1–6. http://dx.doi.org/10.1145/3530982.

Schneider, K., Liskin, O., Paulsen, H., Kauffeld, S., 2015. Media, mood, and meetings. ACM Trans. Comput. Educ. 15, 1–33.

Scott, M.J., Parker, A., McDonald, B., Lewis, G., Powley, E.J., 2019. Nurturing collaboration in an undergraduate computing course with robot-themed team training and team building. In: Proceedings of the 3rd Conference on Computing Education Practice. Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3294016.3294019.

Shaw, M., Tomayko, J.E., 1991. Models for undergraduate project courses in software engineering. In: Proceedings of the SEI Conference on Software Engineering Education. Springer-Verlag, London, UK, pp. 33–71.

Smith, H.H., Smarkusky, D.L., 2005. Competency matrices for peer assessment of individuals in team projects. In: Proceedings of the 6th Conference on Information Technology Education. ACM, New York, NY, pp. 155–162. http://dx.doi.org/10.1145/1095714.1095751.

Sokolowski, E., Oussena, S., 2017. An intervention detection framework for collaborative projects. In: 2017 Intelligent Systems Conference (IntelliSys). pp. 246–251. http://dx.doi.org/10.1109/IntelliSys.2017.8324300.

Soto, C.J., Jackson, J.J., 2013. Five-factor model of personality. Oxf. Bibliogr. Psychol. http://dx.doi.org/10.1093/OBO/9780199828340-0120.

Souza, M., Moreira, R., Figueiredo, E., 2019. Students perception on the use of project-based learning in software engineering education. In: Proceedings of the XXXIII Brazilian Symposium on Software Engineering. ACM, New York, NY, pp. 537–546. http://dx.doi.org/10.1145/3350768.3352457.

Strauss, P., A, U., Young, S., 2011. 'I know the type of people i work well with': Student anxiety in multicultural group projects. Stud. High. Educ. 36, 815–829. http://dx.doi.org/10.1080/03075079.2010.488720.

Tenhunen, S., Männistö, M., Ihantola, P., 2023. A systematic literature review of capstone courses in software engineering. Inf. Softw. Technol. 159, 107191. http://dx.doi.org/10.1016/j.infsof.2023.107191.

Theobald, E.J., Eddy, S.L., Grunspan, D.Z., Wiggins, B.L., Crowe, A.J., 2017. Student perception of group dynamics predicts individual performance: Comfort and equity matter. PLoS One 12, e0181336.

Tomayko, J.E., 1998. Forging a discipline: An outline history of software engineering education. Ann. Softw. Eng. 6, 3–18. http://dx.doi.org/10.1023/A:1018953214201.

Tuckman, B.W., 1965. Developmental sequence in small groups. Psychol. Bull. 63, 384–399.

Wengrowicz, N., Dori, Y.J., Dori, D., 2014. Transactional distance in an undergraduate project-based systems modeling course. Knowl.-Based Syst. 71, 41–51. http://dx.doi.org/10.1016/j.knosys.2014.05.022.

Wieringa, R., Maiden, N., Mead, N., Rolland, C., 2006. Requirements engineering paper classification and evaluation criteria: A proposal and a discussion. Requir. Eng. 11, 102–107. http://dx.doi.org/10.1007/s00766-005-0021-6.

Wilkins, D.E., Lawhead, P.B., 2000. Evaluating individuals in team projects. In: Proceedings of the Thirty-First SIGCSE Technical Symposium on Computer Science Education. ACM, New York, NY, pp. 172–175. http://dx.doi.org/10.1145/330908.331849.

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. ACM, New York, NY, pp. 1–10. http://dx.doi.org/10.1145/2601248.2601268.

Wong, F.M.F., Kan, C.W.Y., Chow, S.K.Y., 2022. From resistance to acceptance in small group work: Students' narratives. Nurse Educ. Today 111, 105317.

Wood, Z.J., Clements, J., Peterson, Z., Janzen, D., Smith, H., Haungs, M., Workman, J., Bellardo, J., DeBruhl, B., 2018. Mixed approaches to cs0: Exploring topic and pedagogy variance after six years of cs0. In: Proceedings of the 49th ACM Technical Symposium on Computer Science Education. Association for Computing Machinery, New York, NY, USA, pp. 20–25. http://dx.doi.org/10.1145/3159450.3159592.

Xiao, L., 2013. The effects of a shared free form rationale space in collaborative learning activities. J. Syst. Softw. 86, 1727–1737. http://dx.doi.org/10.1016/j.jss.2012.07.042.

Yorra, N., 2012. Optimizing Group Learning: A Phenomenological Study Exploring the Experiences of Senior Business Students at a Major Research University (Ph.D. thesis). Northeastern University, Boston, Massachusetts, http://hdl.handle.net/2047/d20002607.

Zvereva, O., Muravyov, A., 2019. IT team building for experiential learning. In: International Conference on Information Science and Communications Technologies. ICISCT, IEEE, pp. 1–4. http://dx.doi.org/10.1109/ICISCT47635.2019.9011963.

Mikko Kokkoniemi is a Ph.D. student at the University of Jyväskylä. He has worked in various project roles in the industry and his research focuses on team issues in CS/SE student projects. He received his bachelor's and master's degree in computer science from the University of Jyväskylä.

Ville Isomöttönen is a Senior Lecturer at the University of Jyväskylä, where he leads a computing education research group. His teaching is located in the Faculty of Information technology. Research interests include multiple perspectives in project-based learning and programming education, while also theoretical and critical education research. He received his Ph.D. from the University of Jyväskylä.