The Role of Voice and Silence Behaviors in Software Development: A Structural Equation Modeling Analysis

Mary Sánchez-Gordón

Ricardo Colomo-Palacios

Alex Sánchez-Gordón

Aliaksandr Hubin

Abstract

Context: Most software companies strive to have high-performing teams and mitigate withdrawal behaviors like being present but unproductive. In this context, psychological safety and developers' perceived impact are suggested as potential drivers of voice and silence behaviors. However, understanding these social aspects of software development entails the incorporation of social science theories.

Objective: This study aims to empirically explore the relationships among software professionals' perceived impact, psychological safety, voice and silence behaviors, burnout particularly withdrawal and performance using a theoretical model.

Method: A survey questionnaire was conducted, resulting in valid responses from software development teams. Then, we analyzed the responses using structural equation modeling (SEM) and a novel semi-confirmatory SEM. All variables were measured using pre-validated instruments.

Results: The findings supported the theoretical model, showing that psychological safety was more related to silence than voice, whereas perceived impact shows a stronger relationship with voice than with silence. Furthermore, silence contributed to higher burnout, whereas voice alleviated it. In contrast, silence showed a negative association with performance, whereas voice was positively associated with it.

Conclusions: This study examines the direct effects of perceived impact and psychological safety on voice and silence behaviors. Additionally, it examines the direct effects of these behaviors on burnout and task performance among software professionals. Consequently, the findings offer both theoretical and practical insights.

1. Introduction

Software development requires a high level of teamwork and communication. According to Lenberg et al. [1], communication is the process by which individuals exchange thoughts, feelings, or information. Communication is also reported recurrently in the literature as a big challenge for software teams [1–4]. In current scenarios, software practitioners across all roles must incessantly communicate with teammates as well as with project stakeholders, including users, analysts, suppliers, customers, and business partners [5]. Recently, implementing communication, coordination, and cooperation strategies within the software teams has been identified as an important approach to mitigate social, people, and process debt [6]. While communication takes different forms, voice behavior is distinct from other types of workplace communication [7] and is considered an asset for enterprises [4,8]. In turn, voice is linked to and intertwined with silence. This means that both voice and silence behaviors can shape a shared understanding of requirements, coordination across software tasks, and timely adaptation to change, while impacting knowledge sharing, innovation, and cohesion.

Despite the large body of work that has been conducted to understand voice and silence behaviors across a wide range of industries, including manufacturing, retail, and healthcare [7], only a few studies exist that have analyzed the effect of those behaviors on software teams. Therefore, there is a limited understanding of these behaviors in SE context, which entails the incorporation of social science theories to prevent the over-rationalization of core phenomena [9]. Although voice and silence are opposites at the level of a single issue, individuals, including software professionals, handle multiple issues at one point in time, as noted by [10]. For example, they may voice certain quality concerns, such as usability and performance, while remaining silent on others, such as reliability and security. In this vein, previous research also highlighted the need for further understanding of whether silence has different predictors than voice [7]. It raises the following research question: Are voice and silence distinct constructs that relate to antecedents and outcomes distinctly in the context of software development?

To address this question, we follow an execution plan presented in our registered report that aims to test a new theory proposed by Sherf et al. [10] about voice and silence at co-located teams, or hybrid software teams (part-remote/part-office arrangements). Sherf et al. propose a conceptual framework for the independence of voice and silence based on the behavioral activation and behavioral inhibition

systems (see details in the background section). This framework proposes that employees' voice and silence (Section 2.1) may have differential relationships with key antecedents, such as perceived impact (2.3) and psychological safety (2.4), as well as outcomes, such as team performance (2.5) and psychological withdrawal (2.6). Empirically demonstrating whether this theory applies to SE context can contribute to the body of knowledge and address the need identified by Lorey et al. [9] of testing social theories in SE context before use.

The remainder of this paper is organized as follows. Section 2 and Section 3 present background and hypotheses. Section 4 presents the execution of the plan presented in our registered report, while Section 5 reports the data analysis and results. In Section 6, we discuss the deviations and limitations of this study and outline the key implications that our study has for the research community. Finally, Section 7 draws some conclusions. A replication package including our (anonymous) dataset, instruments and analysis scripts is available online at [11].

2. Background

In this section, we review the following areas of related work: voice and silence behaviors, behavioral activation and behavioral inhibition systems, perceived impact, psychological safety, work performance psychological safety, and withdrawal.

2.1. Voice and Silence Behaviors

In work settings, employees' silence and voice are two separate concepts. Employees observe inefficient work procedures, aspects to improve, or inappropriate work behaviors on a regular basis [8]. Voice can be defined as chances for personnel to have a say and possibly impact organizational affairs relating to matters that influence their work and the interests of managers and owners [12]. According to Morrison [13], voice is informal, discretionary, and change-oriented. Voice benefits for organizations include improved generation of ideas, better adaptation to changes, or enhanced problem identification [8]. Voice is essential for ensuring that team members' expertise and competence are used properly inside teams [14]. Research has enriched our understanding of how voice may be beneficial for error prevention, organizational functioning, and innovation [15] and how voice affects the voicers and their teams and supervisors [7].

On the employee side, voice mechanisms increase alignment between organizational and personal goals and enhance productivity, commitment and engagement [16,17]. Through resources like voicing ideas, feedback provision, and knowledge sharing, supportive practices and relationships at work foster

psychological safety and influence positive work outcomes such as learning, performance, innovation, and creativity [18]. Although voicing new ideas potentially benefits an organization, it carries certain risks for employees [8]. For instance, the established way of doing things might be challenged by the voicing of new ideas that go against the vested interests of other team members.

On the other hand, silence may reveal circumstances where employees either do not have chances to voice or do not use them for several reasons [19]. For instance, silence is reported as adverse because employees not communicating their ideas and concerns could not only be harmful to psychological and physical health [8] but may also damage organizational interests [20]. So far, much of the research on silence has investigated it in isolation [7] and is scarce [7,8]. As a result of the importance of voice and silence, the decision-making process of employee voice and silence behaviors is worth scholarly inquiry [8].

In SE literature, voice and silence are under-explored constructs. There is a scenario-based experiment [21] that explores three factors that contribute towards team member silence behaviors in a distributed software project: the individual's level of experience, the role of the offending team member, and the individual's responsibility to report. Findings reveal that some of the factors from the silence literature may not be relevant in the context under study. Another experiment [22] investigates how voice and related dimensions of group process (group work and discussion quality) are affected by group size and social presence. Findings show that social presence impacts instrumental and value-expressive voice, as well as related outcomes of group process.

Other SE studies have not focused on or measured voice or silence but have mentioned them. For instance, Greiler et al. [23] identified that developers are more eager to voice and tackle problems to continuously improve developer experience on teams with good psychological safety. In addition, a recent systematic mapping of burnout in SE [24] identified that communication practices within a development team can lead to burnout among software developers. These practices include impolite requests, lack of participation in terms of 'speaking up', openness to criticism, and disagreement among co-workers. Finally, a recent systematic literature review [25] found that whistleblowing, which is another area of literature related to employee silence, remains an understudied area of SE research.

2.2. Behavioral Activation and Behavioral Inhibition Systems

Employee voice is a term that has been used in a variety of contexts and disciplines over time. Despite the advances in the literature on voice and silence [7], similar mediating judgments, e.g., efficacy judgments, can be found in many studies, suggesting that the logic and expected findings could be anticipated in the opposite

direction by using the word “silence” instead of “voice”. From this perspective, silence constitutes a failure to voice whereas voice is a choice not to remain silent [13]. The two constructs are inversely related, such that antecedents promoting one tend to inhibit the other. For instance, leader openness enhances psychological safety, encouraging voice, while reducing silence [7]. However, this view was challenged by Sherf et al. [7,10]. Although they are not the first to advocate for distinguishing voice from silence, uncertainty has persisted regarding the precise nature of their relationship, whether as opposites, as distinct but connected, or as entirely independent. Knoll et al. [26] identified these relationships and how they may manifest in organizations. Sherf et al. further posit that voice and silence are different because they reflect two self-regulatory systems. Voice is an approach behavior that reflects the behavioral activation (approach) system (BAS) whereas silence is an avoidance behavior that reflects the behavioral inhibition (or avoidance) system (BIS). They carried out a meta-analysis and a panel study over six months. As a result, they concluded that the frequency of voice and silence is determined by the extent to which these independent systems are triggered. In our study, we rely on that proposed conceptual framework for the independence of voice and silence.

Previous literature argues that the BAS and BIS are two self-regulatory systems that underlie behavior [27]. The BAS regulates appetitive motive and aims to move toward potential rewards or opportunities for the self, that is, to approach, seek, or achieve them. On the other hand, the BIS regulates aversive motives and aims to move away from potential harm or punishment to the self, that is, to avoid, prevent, or inhibit them. Moreover, different environmental stimuli trigger the BAS and BIS, which are mutually independent. There is also evidence that BAS and BIS are connected to emotional states, with positive or negative emotions occurring when moving towards positive change or positive stimuli (e.g., rewarded) or away from negative change or negative stimuli (e.g., punished).

2.3. Perceived impact

Perceived impact is the extent to which individuals believe that they can influence outcomes at their work [10]. Thus, impact is the opposite to perceiving that the problems arise in the wake of uncontrollability, i.e., "learned helplessness". It is also worth noting that perceived impact and “locus of control” are different since locus of control is a personality trait that remains in all circumstances, whereas the work context influences the perceived impact [28].

2.4. Psychological safety

The concept of psychological safety can be traced back to the 60's, in the early work by Schein and Bennis [29] on organizational change. It gained new attention in the late 90's, largely due to the fundamental contribution made by Edmondson [30]. Since then, psychological safety has been identified as a key factor in understanding phenomena such as team learning, teamwork, voice, and organizational learning [31]. It is positively associated with outcomes such as information sharing, learning behaviors, contentment, and employee engagement [32].

Psychological safety is also an important interpersonal and cognitive concept related to positive outcomes in software teams [33]. Recent studies shows that once psychological safety is reached, agile teams respond with a range of behaviors that improve team dynamics [34] and software quality [35]. Furthermore, psychological safety is an important factor in the frameworks designed to understand and improve the developer experience [36]. However, negative consequences may occur due to excessive psychological safety, which may lead to individuals spending valuable time on trivial matters or losing their motivation to learn.

2.5. Work performance

Previous research on how software developers experience performance suggests centricity around behavioral and social sciences [37,38]. Although several team performance (or effectiveness) models exist in other disciplines, [39] discussed three widely accepted general teamwork effectiveness models and concluded that there is a need to increase the understanding of the genuine characteristics of software development to know what applies. Despite that fact, communication was identified as central in these models. More recently, another study conducted by Strode et al. [40] proposes an agile teamwork effectiveness model (ATEM) for collocated agile development teams. These authors also argue that communication is key to efficient software development in agile teams.

Generic frameworks used broad dimensions to describe work performance. However, individual work performance (IWP) has emerged as a key indicator at both team and organizational levels [41]. Based on conceptual grouping of IWP dimensions found in the literature, three primary categories can be identified: task performance, contextual performance, and counterproductive work behavior. However, the main focus of the work and organizational psychology literature is on task performance [42].

2.6. *Withdrawal*

The burnout phenomenon is a form of occupational fatigue characterized by exhaustion (inability) and withdrawal (unwillingness), according to Schaufeli and Taris [43]. For employees, silence may cause feelings of anger and/or fear, reduce job satisfaction and creativity, and increase burnout [7].

Exhaustion is often subsumed by withdrawal, since individuals deal with exhaustion by using it as a coping mechanism [44]. In turn, employee withdrawal is a set of attitudes and behaviors that is defined as employees' attempts to remove themselves, either temporarily or permanently, from quotidian work tasks perceived by the employees as the source of their negative affect [45]. Theoretically, negative emotions should precede work withdrawal. In the SE context, a previous study [46] found that varying degrees of withdrawal cause developers to distance themselves from the task, which increases unhappiness up to considering quitting developing software or even the point of quitting the job.

3. Hypotheses

The hypotheses are based on established theories and salient themes from previous studies. Fig. 1 illustrates the theoretical model and hypotheses. The theory hypothesizes that the psychological safety and perceived impact will both be related to silence (H1 and H3) and voice professionals (H2 and H4) behaviors among professionals. On the other hand, silence and voice behaviors will influence psychological withdrawal (H5 and H7) and performance (H6 and H8).

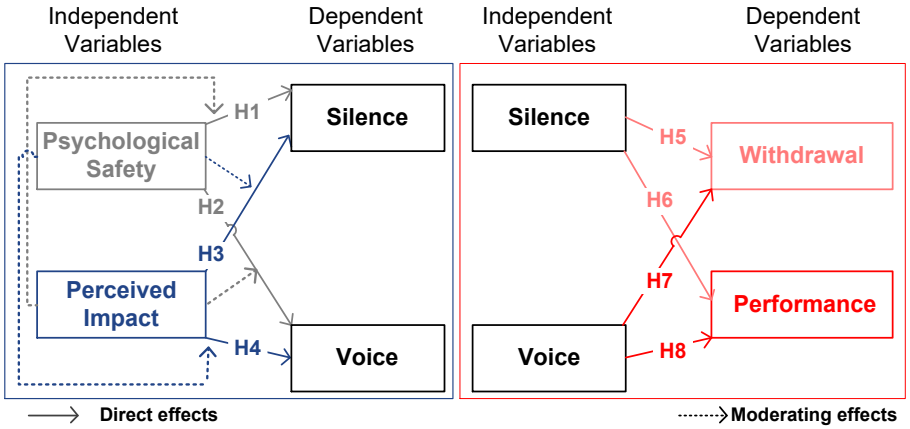

Fig. 1 Theoretical model and the hypotheses

3.1. The Influence of Psychological Safety and Perceived Impact on Voice and Silence Behaviors

Sherf et al. [10] found that "perceived impact and psychological safety differentially predict voice and silence behaviors". Findings showed that voice is more strongly affected by perceived impact (through the BAS activation), but silence is more strongly influenced by psychological safety (by the BIS activation). Our study investigates psychological safety since it is considered a valuable resource [18]. Software industry studies identified psychological safety as a top driver of team and business performance [23]. Additionally, we will investigate a motivational state called perceived impact (influence), which has been found to affect teams in the voice and silence literature, but has been overlooked by SE researchers.

Employees' perceptions of impact seem a proper environmental cue for the BAS since perceived impact is the degree to which employees believe that they can influence outcomes at their work [10]. Perceived impact can alert employees about their ability to gain opportunities or rewards by voicing knowledge or information. For instance, the project team members' perception of the supervisor's openness to the information about an underperforming project's flaws has a positive impact on the willingness to voice concerns about that project to their supervisor [47]. However, in practice, the style of leadership is probably less important than if the employee feels that their manager is supportive, sets a positive example, and creates an environment where speaking up is safe [7]. Employees who perceive greater impact, sense higher levels of control over their work environment and develop a greater sense of responsibility to engage in change-oriented behaviors like voice [48].

High perceived impact levels would increase voice with the objective of achieving desired work outcomes. However, if the environment creates a sense of learned helplessness in which no opportunities to achieve gains or rewards exist, the BAS is less likely to be triggered, leading to low infrequent levels of voice [10]. Approach behaviors like voice are not triggered when there is a low perceived impact because rewards or opportunities are either absent or not achievable. It suggests that silence, as an inhibition-oriented behavior, will be less strongly influenced by perceived impact because it is less associated with BIS states triggered by the presence of risk (e.g., fear, anxiety and vigilance) or by the absence of risk (e.g., security, relief and calmness) [10].

Psychological safety seems a proper environmental cue for the BIS since psychological safety refers to the degree to which behaviors might have negative interpersonal outcomes [7,49]. Psychological safety is a key factor of high-quality communication, decision making, and trust that plays a vital role within workplace teams [18,50]. Psychological safety facilitates interpersonally risky behaviors like

speaking up and voicing [8,18,50]. According to Newman et al. [18], previous work on enhancing psychological safety also suggests that employees feel safer sharing ideas and speaking out when the higher status of the team or employee is. In particular, experienced employees with high work positions are more likely to believe that speaking up will bring them positive outcomes [8]. However, although employees think their contribution would be valuable to the team, they sometimes choose silence over voice when there is a lack of psychological safety [15]. Lack of psychological safety would increase employee silence with the objective of avoiding potential threats, risks or dangers in the work setting. However, if the environment is psychologically safe, the BIS is less likely to be triggered, leading to low or infrequent levels of silence [10]. Avoidance behaviors like silence are not triggered when there is high psychological safety because there will be no negative stimuli (potential harm or punishment). It suggests that voice, as an approach-oriented behavior, will be less strongly influenced by psychological safety because it is less associated with BAS states like hope, enthusiasm, or eagerness [10].

Given that psychological safety and perceived impact can take place simultaneously, e.g., [10,48], we include these predictors as control variables in our hypotheses as recommended by Becker et al. [51] and formulate the following:

Hypotheses H1 and H2: Psychological safety in software development teams relates more strongly to their silence than to their voice when controlling for the effect of team members' perceived impact.

Hypotheses H3 and H4: Software development team members' perceived impact relates more strongly to their voice than to their silence when controlling for the effect of psychological safety.

3.2. Relationships of Voice and Silence with Withdrawal and Performance

Sherf et al. [10] also proposed that silence relates to burnout and provides empirical evidence linking it to withdrawal more strongly than voice. They also suggested that future research should examine the differential relationships of voice and silence with performance.

Prior research [10] suggests that frequent BIS activation can have a negative impact on well-being costs by increasing aversive states like strain, stress, or fatigue. A longitudinal study [52] found reciprocal effects between silence, abusive supervision, and fear. Abusive supervision increased fear and led to more defensive silence, which in turn resulted in more abusive supervision. These effects were intensified by lower assertiveness and higher climate-of-fear perceptions. Moreover, a four-wave longitudinal study [53] found reciprocal effects between

burnout and silence. Two imposed forms of silence (acquiescent and quiescent) were related to increased levels of subsequent burnout, which in turn related to increased silence. Likewise, a survey study [54] revealed that silence leads to burnout, which results in weakening performance, an increase in withdrawal behaviors, and turnover intentions. Through the mediation of job burnout, silence has a more pronounced negative effect on employee performance and a positive impact on turnover intentions. Findings of these studies show that silence has a consistently positive relationship with burnout. From an approach-avoidance perspective, these findings can be understood as a movement away from negative stimuli.

On the other hand, voice effects on occupational well-being have not been well understood [7,55]. Engaging in voice can potentially damage employees' well-being when leading to harmful consequences, e.g., damaged relationships with co-workers and superiors, or loss of career opportunities [55]. It also fosters situations where employees talk about potentially uncomfortable issues and defend alternative points of view, making voice a highly resource-demanding behavior. Moreover, a lack of understanding of how voice may impact well-being is critical since low well-being can result in deterioration of health, withdrawal, and higher turnover [55]. Depending on how (un) favorably others perceive voice, it may either increase or reduce withdrawal and exhaustion [8]. However, research identified that approach motives and sensitivity to BAS have no effect on outcomes similar to burnout, e.g., strain [56]. As a result of this uncertainty, we suggest that voice's relationship with burnout, in particular withdrawal, should be weaker than the relationship between silence and burnout. Our study will examine withdrawal because it has been identified as a relevant factor that affects software development. A recent study [23] identified "speaking up" as a strategy to improve the developer experience, and "no longer speaking up" and "reducing engagement" about the problems as coping mechanisms. The last implies that practitioners still performed only absolutely necessary duties, which means "withholding work efforts" a characteristic of withdrawal behavior. Hence, we propose the following hypothesis:

Hypotheses H5 and H7: Software development team members' silence presents a stronger association with withdrawal compared to their voices.

Research on performance is related to voice (a response to BAS). As BAS is related to obtaining and striving for opportunities and rewards, engaging in voice may affect performance by signaling an employee's prosocial contributions to the team [13]. Recently, a longitudinal approach found that voice more positively impacts team effectiveness when teams face discontinuous change [14]. Particularly, prohibitive voice reduced performance losses by enabling error management at the

discontinuous change phase and promotive voice enhanced performance gains by enabling process innovation at the recovery phase. Moreover, a meta-analysis [57] identified that the relationship between high-performance managerial practices and job performance is mediated by voice. Research also suggests that group and organizational effectiveness suffers when there is a high level of silence and performance is better when employees share their ideas and concerns [13]. As mentioned before, BIS regulates aversive motives and can lead to strain. Silence may have a negative impact on performance by consuming employees' energy and eventually lowering feelings of personal accomplishment [13,53]. Hence, we propose the following hypotheses:

Hypothesis H6 and H8: Software development team members' voice presents a stronger association with work performance compared to their silence.

4. Plan Execution

The plan execution followed our registered report with some deviations, which are mentioned throughout this paper. Specifically, this section outlines our approach, which was guided by the theoretical model shown in Fig. 1.

4.1. Population and Inclusion Criteria

Initially, we targeted software development teams, aiming to recruit all team members through a single global campaign that employed an English-language questionnaire as outlined in our registered report. While we anticipated that response rates would be a challenge in studies like this, and took steps to minimize the time commitment for respondents, our preliminary analysis of the data collected between June and November 2023 indicated that this approach was ineffective. The low response rate was not only due to difficulties in obtaining responses from all team members, but also because the proportion of valid responses was much lower than expected.

Consequently, we decided to revise our convenience sampling strategy by shifting the focus from team to individual software professionals and conducting local campaigns to recruit participants from Spanish speaking countries. Although the initial data collection was unsuccessful, we decided to maintain our original plan from the registered report and did not offer cash incentives for participation. Furthermore, as originally planned, the questionnaire was made available to a broad range of software professionals, including designers, project managers, quality assurance specialists, architects, and business analysts.

4.2. Instrument Design

The questionnaire was designed alongside the hypotheses prior to initiating the data collection process. In line with the registered report, we developed an anonymous questionnaire using the survey tool QuestionPro. Therefore, we avoided using URL tracking or collecting any contact information.

Questions were grouped into three main parts i) demographic information, ii) professional experience, and iii) theoretical model. Fig. 2 provides an overview of the questionnaire, while the complete survey is available in the replication package [11].

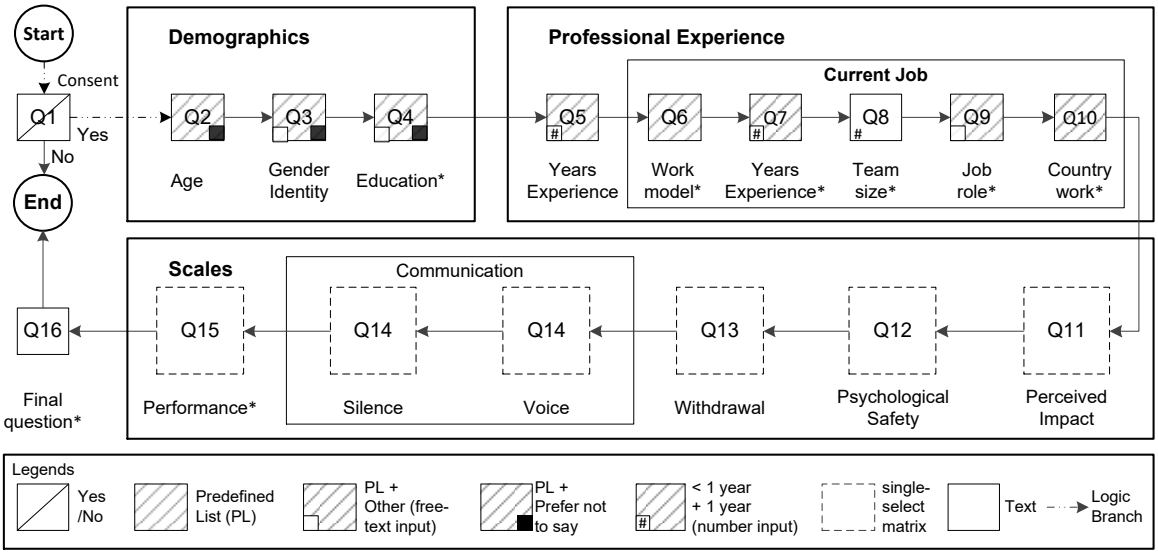

Fig. 2 Overview of the structure of the questionnaire with an asterisk (*) indicating the differences from the registered report.

The first two sections of the questionnaire featured single-choice selection questions based on predefined lists (PL). The first section focused on basic demographic information including participant age (Q2; PL: 18-25, 26-35, 36-45, 46-55, 56-65, 66+ years), gender identity (Q3; PL: man and woman), and education (Q4; PL: undergraduate degree, master's degree, and doctoral degree). These questions provided the option “Prefer not to say,” and the last two included free-text input fields: “I prefer to describe myself” for gender identity and “Other” for education. However, education was not part of our registered report. The second section addressed participants’ years of experience (Q6). Different from the registered report, this section included questions related to participants’ current job: work model (Q6; PL: Remote, Hybrid, and On-site), years of experience (Q7), team size (Q8), job role (Q9; PL: analyst, architect, developer, tester, project manager), and country where the participant works (Q10).

The third section of the questionnaire included single-select matrix questions based on the following validated scales to strengthen construct validity.

Perceived impact (Q11; Pi) is the extent to which individuals believe that they can influence outcomes at their work [10]. To capture respondents' perceived impact on their teams, we used a three-item measure (1 = "strongly disagree" to 5 = "strongly agree") [28].

Psychological safety (Q12; Ps) is defined as "a shared belief held by members of a team that the team is safe for interpersonal risk-taking" [30]. For the BIS is relevant and highly salient that perceptions about the possibility that one's communication will result in a negative outcomes such as discomfort or rejection [10]. To capture individuals' experience of psychological safety in their team, we used a seven-item measure developed in [30]. Items were each measured on a seven-point Likert scale (1 = "strongly disagree" to 7 = "strongly agree").

Withdrawal (Q13; Bo). Withdrawal manifests through a set of attitudes and behaviors such as increased resistance, reduced commitment, lack of interest, disengagement, mental distancing. To assess psychological withdrawal, we use eight items from [58]. All items are scored on a 7-point Likert scale (1 = "never" to 7 = "daily").

Voice (Q14; Vo) is the expression of concerns and ideas with the goal of influencing others in order to induce improvements or changes [10]. Workplace response to BAS regulation is voicing issues to obtain successes, rewards, or opportunities [10]. In the last decade, according to Morrison [7], there has been increasing use of the measure proposed by Liang et al. [49]. To capture respondents' voice, we utilized a promotive (three items) and prohibitive (three items) voice measure (1 = "Never" to 7 = "Daily") from Liang et al. [49] and adapted by Sherf et al. [10].

Silence (Q14; Si) is the suppression of concerns and ideas with the goal of preventing or avoiding communication that could cause harm to oneself [10]. Workplace response to BIS regulation is withholding issues to avoid failures, risks, or punishments. According to Morrison [7], the general, non-motive-based measures proposed by Tangirala and Ramanujam [48] and Detert and Edmondson [15] are among the most widely used. To capture respondents' silence, we utilize a three-item silence scales (1 = "Never" to 7 = "Daily") adapted from those measures by Sherf et al. [10].

Work performance (Q15; Wp) is defined as a "proficiency or ability to perform the tasks that are central to one's job" [42]. This dimension contributes to the organization's technical core by including ability in planning and organizing work, quality of work, results-oriented, and ability to work efficiently. To assess work

performance, we use five items (1 = “rarely” to 5 = “always” defined in [59] based on the items suggested in [41].

Due to the shift in focus from teams to individual software professionals, we selected a different scale for work performance (Q15) than the one defined in our registered report. Additionally, we included the following final open-ended question (Q16): “Is there anything else you want to tell us about how team members' behaviors affect your software development team?”

Finally, the English questionnaire was first translated into Spanish by a bilingual native Spanish speaker (forward translation). A bilingual native English speaker then translated it back into English (backward translation). We compared both versions to identify any discrepancies or inconsistencies. For the work performance scale, the Spanish version was also cross-checked against a previously validated Spanish version [59]. As a result, no issues were found.

4.3. Pilot

We asked five colleagues, including two software engineering academics and three software developers, for feedback on the questionnaire. Their comments focused on the structure, as well as the face and content validity of the scales. Based on this feedback, we made some minor changes, such as moving the Performance scale closer to the end, adding an open-response question, and including the participant's country of work. These changes helped us enhance the content validity of the questionnaire.

4.4. Data Collection

After the pilot, the first author recruited collaborators based in Mexico and Ecuador to assist in disseminating the survey, while the other authors, located in Panama and Spain, contributed to this effort. This approach ensured that participants were software professionals and leveraged each collaborator’s local knowledge to reach a broader pool of potential participants in each country. Nevertheless, the questionnaire included a filter question to exclude participants who were not software professionals and/or who did not provide consent. denied consent, redirecting them to the end of the questionnaire.

Participants received an invitation with a survey link and were encouraged to share it with colleagues who might also be interested. As the survey was anonymous, we did not request participants to provide their colleagues' email addresses. A remainer was sent one-week later. The survey was available from November 2024 to January 2025.

5. Analysis and Results

We received a total of 215 responses of which 51 were incomplete. As a result, we obtained 164 complete responses. This section describes how the data was cleaned and analyzed. The dataset is available in the replication package [11].

5.1. Data Cleaning

After exporting the data to an Excel file, the data was cleaned as follows. To preserve anonymity, we delete the timestamp field. Then, we examine the consent form confirmation field (checkbox, response YES/NO). As we expected, all respondents consented to participate, and this field was deleted since the response was always "YES". As mentioned before, a total of 215 responses were received but 51 were incomplete. Among the 164 participants who provided complete responses, six reported not being software developers, leaving 158 valid responses.

To calculate the value for items on a reversed scale, the following formula is used: reverse scale(x) = max(x) + 1 - x. This adjustment was made for three items in the Psychological Safety Scale: Ps-1 "If I make a mistake on this team, it is often held against me", Ps- 3 "People on my team sometimes reject others for being different", and Ps- 5 "It is difficult to ask other members of my team for help". For example, on a 7-point Likert scale, if the original score is 1 and the maximum value is 7, the calculation would be (7 + 1 = 8) and then subtract the original score from the result (8 - 1 = 7).

5.2. Demographics

All 158 participants self-identified as current software professionals. Table 1 shows an overview of the frequency of responses. The majority of participants identified as men (n = 123, 77.85%), while 33 (20.89) identified as women and two preferred not to disclose their gender (1.27%). Although most participants (91.14%, 144) were under 55, 14 (8.86%) preferred not to disclose their age. Moreover, most participants indicated having an undergraduate degree (86, 54.43%), followed by a graduate degree (56, 35.44%), while the remaining 16 indicated holding other types of degrees.

Additionally, the highest share of participants reported working on-site (38.61%) although remote (37.34%) and hybrid (24.05%) models were also commonly reported. As we expected, the largest job role among participants were developers (52.53%, 83). Due to the sampling strategy, the participants were working in Ecuador (42.41%, 67), Spain (22.15%, 35), Panama (2.15%, 35), Mexico (10.76%, 17), and the United States (2.53%, 4). Regarding total years of professional experience, 13.92%

of participants (22 individuals) had less than one year, while 86.08% (136 individuals) had more than one year, with an average of 9.3 years. In their current job, 23.42% (37) had less than one year of experience, while 76.58% (121) had more than one year, with an average of 6.3 years in the current role.

Table 1 Overview of frequency of responses (participants, n= 158)

Demographic information	**#**	**%**	**Professional experience**	**#**	**%**
Gender			**Working model**		
Male	**123**	**77.85**	Remote	59	37.34
Female	33	20.89	Hybrid	38	24.05
Prefer not to say	2	1.27	On-site	**61**	**38.61**
Age			**Job role**		
Under 26	25	15.82	Analyst	7	4.43
26-30	29	18.35	Architect	10	6.33
31-35	33	20.89	Developer	**83**	**52.53**
36-45	**41**	**25.95**	Tester	16	10.13
46-55	16	10.13	Project Manager	13	8.23
Prefer not to say	14	8.86	Others	29	18.35
Education			**Country**		
Undergraduate	**86**	**54.43**	Ecuador (52)	**67**	**42.41**
Graduate	56	35.44	Mexico (114)	17	10.76
Doctorate	3	1.90	Panama (135)	35	22.15
Prefer not to say	2	1.27	Spain (163)	35	22.15
Other	11	6.96	US (187)	4	2.53
Years of experience (Total)			**Years of experience (Current Job)**		
Less than 1 year	22	13.92	Less than 1 year	37	23.42
More than 1 year (mean=9.3)	**136**	**86.08**	More than 1 year (mean=6.3)	**121**	**76.58**

5.3. Evaluation of the Measurement Model

First, we assessed content validity through a pilot study (Section 4.3). Then, statistical analysis was conducted. The scree plot illustrated in Fig. 3 indicates that six latent factors are present in our data.

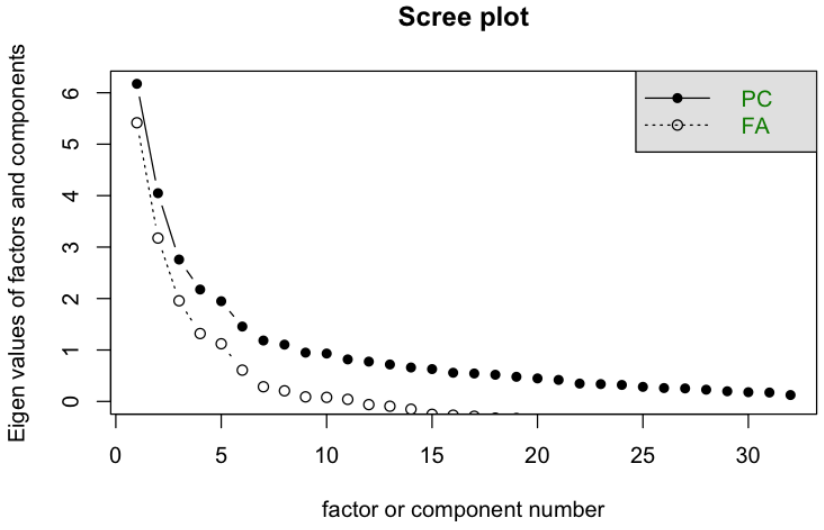

Fig. 3 Scree plot, indicating six factors or components before a linear trend.

Hence, exploratory factor analysis (EFA) with 6 factors was conducted on the survey items by the principal factor method with Varimax rotation that determined the dimensions of each variable examined. Table 2 displays the results of EFA that delineated each variable into a specific group.

Table 2 Results of exploratory factor analysis. Notes: * p< 0.05; **p < 0.01. A cutoff of 0.4 was used for factor loadings.

Construct	Bartlett's test of sphericity and df	KMO	Construct (% explained of total variance)
Perceived impact (Pi)	**227.219****	**0.735**	**Pi (15.99%)**
Pi1: My impact on what happens in my team is large			0.804
Pi2: I have a great deal of control over what happens in my team			0.752
Pi3: I have significant influence over what happens in my team			0.828
Psychological safety (Ps)	**175.306****	**0.734**	**Ps (13.81%)**
Ps1: If I make a mistake on this team, it is often held against me			0.631
Ps2: Members of my team are able to bring up problems and tough issues			-
Ps3: People on my team sometimes reject others for being different			0.675
Ps4: It is safe to take a risk around here			0.402
Ps5: It is difficult to ask other members of my team for help			0.585
Ps6: No one on my team would deliberately act in a way that undermines my efforts			-
Ps7: Working with my team members, my unique skills and talents are valued and utilized			0.473
Voice (Vo)	**464.361****	**0.819**	**Vo (21.11%)**
Vo1: I proactively gave suggestions for issues that may influence my team			0.740
Vo2: I proactively voiced constructive suggestions which are beneficial to my team			0.728
Vo3: I made suggestions on how to improve my team's working procedures			0.774
Vo4: I advised against undesirable behaviors that would hamper my team's job performance			0.504
Vo5: I spoke up honestly about problems that might cause serious loss to the work team, even when/though dissenting opinions exist			0.773
Vo6: I pointed out problems when they appeared in my team, even if that would hamper relationships with other colleagues			0.700
Silence (Si)	**139.507****	**0.698**	**Si (11.26%)**
Vo7: I kept quiet and did not make recommendations about how to fix work-related problems			0.614
Vo8: I kept ideas about how to improve work practices to myself			0.811
Vo9: I choose not to speak up with ideas for new or more effective work practices			0.711
Withdrawal (Bo)	**357.366****	**0.822**	**Bo (18.34%)**
Bo1: I think of being absent			0.800
Bo2: I chat with co-workers about nonwork topics			-
Bo3: I leave my work station/area for unnecessary reasons			0.771
Bo4: I daydreaming			0.599
Bo5: I spend work time on personal matters			0.735
Bo6: I put less effort into my work than I should have			0.714
Bo7: I think of leaving my current job			0.643
Bo8: I let others do my work			0.594

Work performance (Wp)	**324.969****	**0.845**	**Wp (19.49%)**
Wp1: I managed to plan my work so that I finished it on time			0.688
Wp2: I kept in mind the work result I needed to achieve			0.666
Wp3: I was able to set priorities			0.790
Wp4: I was able to carry out my work efficiently			0.714
Wp5: I managed my time well			0.736

The results using a cutoff of 0.4 for factor loadings suggest that Perceived impact (Pi), Psychological safety (Ps), Voice (Vo), Silence (Si), Withdrawal (Bo) and Work performance (Wp) were loaded on one factor, comprised of three, five, six, three, seven and five items, respectively, with 15.99%, 13.81%, 21.11%, 11.26%, 18.34%, and 19.49% of the total variance explained. The factor loadings of the items ranged from 0.402 to 0.811 for factors above the cutoff value of 0.40. Bartlett's tests of sphericity for each factor with values of 227.219, 175.306, 357.366, 464.361, 139.507, and 324.969, respectively, were all strongly significant ($p<0.001$). Moreover, Kaiser–Meyer–Olkin (KMO) statistics of sampling adequacy are high (KMO of 0.735, 0.734, 0.822, 0.819, 0.698, and 0.845, respectively). However, some loadings of Withdrawal (Bo) and Psychological safety (Ps) with a cutoff below 0.2 were initially loaded on two and three factors. To ensure a single-factor structure, Bo2, Ps2 and Ps6 were removed. Previous research also found convergent validity issues with the psychological safety scale and removed Ps2: "I have a great deal of control over what happens in my team" [34]. The resulting EFA allowed to increase the total variance explained from 49.35% to 53.09%, with respectively 15.15%, 13.43%, 21.33%, 11.58%, 18.62%, and 19.88% of total variance explained by the factors. For reduced factors, corresponding to Withdrawal (Bo) and Psychological safety (Ps), the Bartlett's tests remained strongly significant with test statistics of 147.872 and 338.846, and KMO values of 0.730 and 0.826 respectively. We further proceed with internal consistency analysis for the reduced factors.

Internal Consistency. This check aims to ensure that the items consistently and reliably measure the intended latent variables. Consequently, we assessed the reliability of the scales by using a Cronbach's alpha (α). Table 3 shows that the scales used for the variables demonstrated good reliability (Cronbach's $\alpha > .70$) overall. We can conclude that our tests meet the reliability criteria. Pearson's bivariate correlation analyses demonstrated that perceived impact ($r = -0.263$) and psychological safety ($r = -0.254$) had statistically significant negative relationships with professionals' withdrawal, while only perceived impact ($r = 0.216$) had statistically significant positive relationships with professionals' performance. Voice showed positive associations with perceived impact ($r = 0.302$) while silence showed negative associations with psychological safety ($r = -0.232$). As expected, silence was positively associated with Withdrawal ($r = 0.383$) and negatively associated with performance.

Table 3 Cronbach's α, means, standard deviation (SD) and correlations for key study variables. Notes: * p< 0.05; **p < 0.01

	Scales	Items	α	M	SD	1	2	3	4	5	6
1	Pi	3	0.867	3.68	0.928	1	.298**	.302**	-0.094	.216**	-.263**
2	Ps	5	0.706	5.65	1.036		1	0.034	-.232**	0.151	-.254**
3	Vo	6	0.856	4.64	1.317			1	-0.028	0.148	0.026
4	Si	3	0.788	2.73	1.478				1	-0.098	.383**
5	Wp	5	0.855	4.45	0.525					1	-.344**
6	Bo	7	0.809	2.11	0.993						1

5.4. Evaluation of the Structural Model

To test our hypotheses, we use structural equation modeling (SEM) since this study aims to test a theory involving constructs, i.e., latent variables. In this study, Perceived impact (Pi), Psychological safety (Ps), Voice (Vo), Silence (Si), Withdrawal (Bo) and Work performance (Wp) act as the constructs.

To design a structural equation model, we first define a measurement model, which maps each reflective indicator into its corresponding construct. For example, each of the three items comprising the Perceived impact (Pi) scale is modeled as a reflective indicator of perceived impact. The causal SEM structure then follows Fig.1. SEM then uses confirmatory factor analysis (CFA) to estimate each construct as the shared variance of its respective indicators while also estimating the strength of each relationship and the overall accuracy of the model.

Thus, CFA was conducted to verify that the measurement model is consistent. Here, the latent concepts are captured by their respective exogenous variables. A robust WLSMV estimator was used to account for the data observed on a Likert scale. We evaluated the model with all variables included in the factors (m_1), and the model with reduced factors for Withdrawal and Psychological Safety (m_2). Both models were executed, and all diagnostics were run. We evaluate model fit by inspecting several indicators, as shown in Table 4. Here, we see that both m_1 and m_2 have significant differences from the respective saturated models and have many of the other fit criteria slightly worse than the desired cutoffs.

Table 4 Model fit assessment based on recommended indicators [60]. Notes: †value does not satisfy fit criteria

Indicators	m_1	m_2	m_3	Cutoff Value
Chi-Squared	500.718	653.210	144.412	Lower values indicate a better fit. Ideally, a non-significant chi-square (p>0.05) is desired
Chi-Squared / df ratio	1.364	1.436	0.806	Ratios less than 3 are considered indicative of a good fit

Indicators	m_1	m_2	m_3	Cutoff Value
P-value	**0.000**†	**0.000**†	0.973	P-value of Chi-Squared test with respect to the saturated model (above 0.05 is desired)
Comparative Fit Index (CFI)	**0.943**†	**0.920**†	1.000	Values above 0.95 suggest a good comparative fit between the hypothesized model and the observed data
Tucker-Lewis Index (TLI)	**0.937**†	**0.913**†	1.037	Similar to CFI, values above 0.95 are generally considered good.
Root Mean Square Error of Approximation (RMSEA)	0.048	0.053	<0.001	Values less than 0.06 indicate a good fit
Standardized Root Mean Square Residual (SRMR)	**0.089**†	**0.093**†	0.065	Values less than 0.08 are generally considered favorable

Further to check robustness of our conclusions from models m_1 and m_2 to the model that satisfy all assumptions regarding the fit, we used the optimal semi-confirmatory model (m_3) with respect to including or not including each of the observed variables in their respective factors, found by parallel Mode Jumping Markov Chain Monte Carlo (MJMCMC) search [61] implemented in FBMS (version 1.1) R package . This approach is particularly suited for high-dimensional discrete spaces with multimodal likelihood surfaces, such as those arising in semi-confirmatory SEM.

For the MJMCMC, 16 parallel chains for 250000 iterations each were run. The Bayesian Information Criterion (BIC) allowing for computing the approximate marginal likelihood of each model m of exp(-0.5×BIC_m) is not available for WLSMV estimator due to its non-likelihood based nature, we did not target marginal model posteriors. Instead, we defined an auxiliary target distribution of form (1):

$$q(m) \propto \exp(CFI\_m + TLI\_m - RMSE\_m - SRMR\_m) \quad (1)$$

Here, CFI_m is the comparative fit index of the model m, TLI_m is the Tucker-Lewis Index of model m, RMSE_m is the root mean squared error of approximation, and SRMR_m is the standardized root mean square residual of model m. We first ran unconstrained optimization, which resulted in a perfect fit but reduced the factors to being not satisfying validity and internal consistency requirements. Hence, we reformulated the optimization into a constrained integer optimization problem, where we force the KMO and Cronbach α of each factor in the considered models being above 0.6, and Bartlett's tests results being significant on 0.05 level. These constraints additionally induce regularization and prevent overfitting. The final optimization problem is defined in Equation (2):

$$\max_{\mathrm{m}} \{\exp(CFI_m + TLI_m - RMSE_m - SRMR_m)\}, \text{where}$$

$$\mathrm{m} = \left(\gamma_1^{\mathrm{Pi}}, \gamma_2^{\mathrm{Pi}}, \gamma_3^{\mathrm{Pi}}, \gamma_1^{\mathrm{Ps}}, \ldots, \gamma_7^{\mathrm{Ps}}, \gamma_1^{\mathrm{Vo}}, \ldots, \gamma_6^{\mathrm{Vo}}, \gamma_1^{\mathrm{Si}}, \gamma_2^{\mathrm{Si}}, \gamma_3^{\mathrm{Si}}, \gamma_1^{\mathrm{Bo}}, \ldots, \gamma_8^{\mathrm{Bo}}, \gamma_1^{\mathrm{Wp}}, \ldots, \gamma_5^{\mathrm{Wp}}\right), \gamma_i^K \in \{0,1\},$$

subject to

$$\mathrm{KMO}_{\mathrm{m}}^{\mathrm{Pi}} > 0.6\text{ , and } \alpha_{\mathrm{m}}^{\mathrm{Pi}} > 0.6, \text{and } \mathrm{bp}_{\mathrm{m}}^{\mathrm{Pi}} < 0.05;$$

$$\mathrm{KMO}_{\mathrm{m}}^{\mathrm{Ps}} > 0.6\text{ , and } \alpha_{\mathrm{m}}^{\mathrm{Ps}} > 0.6, \text{and } \mathrm{bp}_{\mathrm{m}}^{\mathrm{Ps}} < 0.05;$$

$$\mathrm{KMO}_{\mathrm{m}}^{\mathrm{Vo}} > 0.6\text{ , and } \alpha_{\mathrm{m}}^{\mathrm{Vo}} > 0.6, \text{and } \mathrm{bp}_{\mathrm{m}}^{\mathrm{Vo}} < 0.05;$$

$$\mathrm{KMO}_{\mathrm{m}}^{\mathrm{Si}} > 0.6\text{ , and } \alpha_{\mathrm{m}}^{\mathrm{Si}} > 0.6, \text{and } \mathrm{bp}_{\mathrm{m}}^{\mathrm{Si}} < 0.05;$$

$$\mathrm{KMO}_{\mathrm{m}}^{\mathrm{Bo}} > 0.6\text{ , and } \alpha_{\mathrm{m}}^{\mathrm{Bo}} > 0.6, \text{and } \mathrm{bp}_{\mathrm{m}}^{\mathrm{Bo}} < 0.05;$$

$$\mathrm{KMO}_{\mathrm{m}}^{\mathrm{Wp}} > 0.6\text{ , and } \alpha_{\mathrm{m}}^{\mathrm{Wp}} > 0.6, \text{and } \mathrm{bp}_{\mathrm{m}}^{\mathrm{Wp}} < 0.05$$

The MJMCMC search procedure used to solve Equation (2) is outlined in the replication package [11] (see Supplementary_Appendix_A). The resulting optimal model, m_3, excluded Ps2 and Ps7 from the Psychological Safety factor. It retained only Bo3, Bo4, and Bo5 within the Withdrawal factor. Additionally, it removed Wp3 variable from Work Performance factor and excluded Vo4, Vo5, Vo6 from the voice factor. The remaining factors are identical to m_1 and m_2. The three models m_1,m_2,m_3 had the fit statistics reported in Table 4. As expected, m_3 satisfies all fit criteria well in the strictest sense. While m_1 and m_2 were significantly worse than the respective saturated (p-values of Chi-Squared test <0.001) models and had SRMR above 0.08.

Although we obtained a better SEM fit for m_3, the underlying factors obtained from a semi-confirmatory model need to be studied. By design, m_3 satisfies all psychometric constraints, including factor reliability and adequacy and thus has satisfactory validity and internal consistency. In Table 5, we report the characteristics of the factors in m_3.

Table 5 Characteristics of underlying latent factors for m_3.

	Factors	Items	Bartlett's statistics	df	p-value	KMO	α
Pi	Perceived Impact	3	227.219	3	<0.000	0.735	0.867
Ps	Psychological Safety	5	116.100	10	<0.000	0.706	0.617
Vo	Voice	3	227.966	3	<0.000	0.717	0.859
Si	Silence	3	139.507	3	<0.000	0.698	0.788
Bo	Withdrawal	3	94.641	3	<0.000	0.644	0.683
Wp	Work Performance	4	205.780	6	<0.000	0.769	0.810

Once the model m_3 was executed and all diagnostics passed, that is, lavaan ended normally after 75 iterations with 52 free parameters and n = 158. We evaluate model fit by inspecting the indicators listed in Table 4.

The CFA fit by a robust WLSMV estimator converged in the model m_3 and all the indicators load well on their constructs as shown in Fig. 3. The estimates for Ps1,

Ps3, and Ps5 are positive because these items were reversed (i.e. lower score = better psychological safety).

Table 6 Structural equation model regressions for m_3. Notes: * p< 0.05; **p < 0.01

Regressions		Estimate	Std.Err	z-value	P(>|z|)	Std.lv
Voice~	Psychological Safety	-0.038	0.169	-0.224	0.823	-0.025
	Perceived Impact	0.546	0.141	3.884	<0.001**	0.491
Silence~	Psychological Safety	-0.473	0.213	-2.222	0.026**	-0.271
	Perceived Impact	-0.096	0.14	-0.688	0.491	-0.076
Work Performance~	Voice	0.146	0.059	2.483	0.013**	0.299
	Silence	-0.064	0.044	-1.443	0.149	-0.149
Withdrawal~	Voice	-0.116	0.105	-1.111	0.267	-0.103
	Silence	0.356	0.11	3.223	0.001**	0.36

Table 6 provides the central to our hypotheses estimates for model m_3, while in Table 7, we summarize the p-value for the three SEM models addressed. Furthermore, Table 7 shows the estimates and interpretations of hypotheses H1-H8 based on the model m_3, along with the corresponding conclusions.

Table 7 Hypotheses assessment. Notes: † factor loading issues; * p< 0.05; **p < 0.01

Hypotheses	m_1	m_2	m_3			
	p-value	p-value	p-value	estimate	Interpretation	Conclusions
H1	**0.066**†	0.035*	0.026*	-0.473	Accepted	Psychological safety is inversely related to both the software developer's silence (H1) and voice (H2), but it is significantly and more strongly associated with changes in his/her silence.
H2	0.496	0.442	0.823	-0.038	Rejected	
H3	<0.001**	<0.001**	<0.001**	0.546	Accepted	Change in the software developer's voice is directly and significantly related to perceived impact (H3) showing a stronger association than his/her silence (H4), which is negatively related but not significant.
H4	0.502	0.315	0.491	-0.096	Rejected	
H5	0.001**	0.001**	0.001**	0.356	Accepted	Change in the software developer's silence is directly and significantly related to withdrawal (H5) showing a stronger association than his/her voice (H7), which is negatively related but not significant.
H7	0.664	0.791	0.267	-0.116	Rejected	
H6	**0.052**†	**0.052**†	0.013*	0.146	Accepted	Work performance is directly and significantly related to the software developer's voice (H6) whereas silence (H8) is negatively related but not significant.
H8	0.013*	0.020*	0.149	-0.064	Rejected	

In summary, all diagnostics indicate the model m_3 is safe to interpret (i.e. CFI = 1.000, RMSEA < 0.001, SRMR = 0.065). Fig. 4 illustrates the supported structural equation models. The arrows between the constructs show the supported relationships. The path coefficients (the numbers on the arrows) indicate the relative strength and direction of the relationships. For example, the arrow from Psychological Safety to Silence indicates that the hypothesis that Psychological Safety affects Silence was supported. The label (−0.27) indicates an inverse relationship (less Psychological Safety leads to more Silence) and 0.27 indicates the strength of the relationship.

The findings are supported by model m_3. Hypotheses H1, H3, H5, and H6 are accepted while Hypotheses H2, H4, H7, and H8 are rejected. That is, psychological safety is inversely related to both silence (H1) and voice (H2), but it is significantly and more strongly associated with changes in silence; change in voice is directly and significantly related to perceived impact (H3) unlike silence (H4). Additionally, silence is directly and significantly related to withdrawal (H5) showing a stronger association than voice (H7), which is negatively related but not significant. Similarly, change in work performance is directly and significantly affected by voice (H6) whereas silence (H8) is negatively related but not significant.

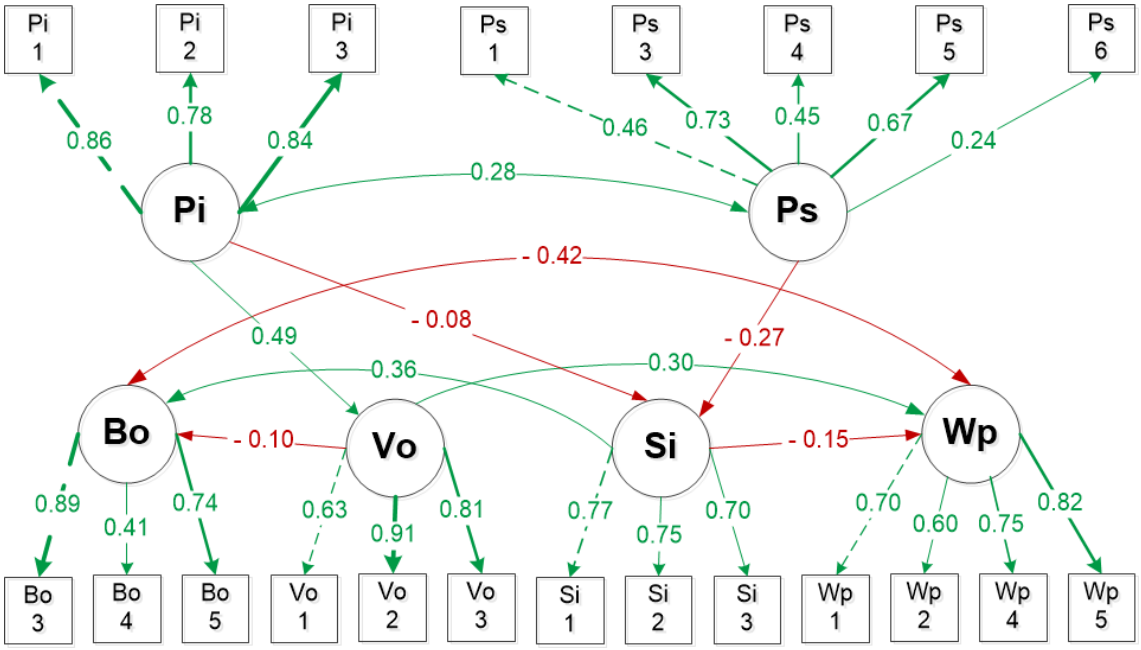

Fig. 4 Full SEM for m_3.

5.5. Exploratory Findings

Inspecting the detailed SEM results (Table 6 6) reveals interesting patterns. Direct relationships include:

Psychological safety in software development teams is not associated with voice.

Software developers' perceived impact is associated with an increase in voice.

Psychological safety in software development teams is associated with a reduction in silence.

Software developers' perceived impact is not associated with silence.

Software developers' voice is associated with an increase in work performance.

Software developers' silence is not associated with work performance.

Software developers' voice is not associated with withdrawal.

Software developers' silence is associated with an increase in withdrawal.

Possible indirect relationships of perceived impact and psychological safety on withdrawal and performance were also examined. However, none of the indirect relationships appeared significant.

In addition, we omitted education, job role, working model and country because SEM does not respond well to nominal categorical variables. Therefore, more research is needed to understand the nature of these effects.

5.6. Summary Interpretation

This study shows that psychological safety and perceived impact reflect relevant environmental cues. Psychological safety triggers the BIS, prompting individuals to avoid self-relevant harm through silence, whereas perceived impact can activate the BAS, encouraging the pursuit of self-relevant rewards and opportunities through voice. Furthermore, frequent silence contributes more to withdrawal-related experiences than voice mitigates them. Conversely, frequent voice has a more positive association with performance than silence has a negative one.

6. Discussion

This study contributes to the scarce research on voice and silence in the SE context by examining their distinction through the lens of BAS and BIS as proposed by Sherf et al. [10]. Our findings challenge common understandings of the underlying psychological mechanisms driving voice and silence behaviors. In doing so, we bring a new perspective on these behaviors, which have not been previously explored in the SE context.

Our findings show that silence is strongly and significantly negatively associated with psychological safety while voice is positive but not significantly related. This contrasts with Sherf et al.'s finding that psychological safety negatively affects voice when perceived impact is controlled. Our measurement model also demonstrated adequate fit to the data [$\chi 2$ (70) = 144.412, CFI = 1.000, TLI = 1.037, SRMR =0.065, and RMSEA < 0.001]. Despite this difference, both Sherf et al.'s findings and ours consistently indicated that a lack of psychological safety activates BIS, leading to increased silence among professionals, suggesting that psychological safety plays

a role in shaping silence behaviour. We also included a freely estimated covariance between voice and silence that had a negligible effect on the main substantive conclusions, suggesting that modeling them as independent is not inconsistent with the observed data. Consistent with Sherf et al.'s [10] study conducted with U.S. employees, our findings show that both voice and silence are associated with withdrawal but in opposite directions, with silence having a significant association. Moreover, our findings provide empirical evidence that supports Sherf et al.'s [10] theoretical extensions on the relationships among voice, silence and performance. As expected, voice, an approach-oriented behavior, is positively and significantly associated with performance, showing a stronger association than silence, an avoidance-oriented behavior that is negatively related to performance.

Possible indirect relationships appeared nonsignificant in our data. Previous research also remains inconclusive on the relationship between psychological safety and performance. While some argue for a direct effect [38,62], others suggest an indirect effect mediated by factors such as team learning, team efficacy [63], team capabilities, and customer involvement [64]. This suggests that performance is influenced by complex factors, emphasizing the need for larger studies that analyze their effects in more detail.

6.1. Practical Implications

Despite the importance of speaking up, software professionals are often reluctant to voice their concerns and choose to remain silent [21,65,66]. Our findings offer insight into this behavior: silence is primarily driven by a lack of psychological safety, whereas voice is more strongly driven by a sense of perceived impact. Researchers and practitioners should consider that professionals who frequently speak up can be withholding other issues. Likewise, a lack of voice should not be interpreted as intentional withholding of concerns. Therefore, directly examining silence motives would offer insights into situations where software professionals have meaningful contributions to make but choose to remain silent.

From this perspective, software professionals perceive a lack of psychological safety as a signal related to BIS and tend to avoid situations that may result in punishment or negative experiences. For example, this may include developers avoiding blame for errors during agile retrospectives [66], disagreeing with suggestions such as reducing the deadline for a project, fearing the consequences of sharing negative feedback or avoiding seeking help [67]. In this context, no blame and collective decision-making practices remove interpersonal risks to enhance psychological safety, which in turn reduces team members' silence [34]. In agile teams, openness in communication is another element of psychological safety that should be actively cultivated [66], particularly as team members may feel

vulnerable when receiving feedback during stand-ups, retrospectives, or issue tracker discussions [34]. A proactive leadership approach is also required [68], Moreover, frequent silence generates experiences of withdrawal more than voice reduces them. Consequently, initiatives for improved professionals' well-being at work may need to focus on addressing silence, rather than voice.

When software professionals perceive they can influence outcomes at work, it triggers BAS. However, fostering frequent voice behaviors requires generating BAS experiences by providing opportunities for success, rewards, or meaningful impact. It suggests that motivation in terms of recognition and appreciation is an important factor, particularly in open-source development [69], where many of the developers are volunteers. Furthermore, marginalized software professionals like women [70] face greater challenges in having their voices heard. Although mentorship and sponsorship programs can improve their experience, they are more willing to voice concerns or ideas when managerial practices are fair [71]. Indeed, both developers and managers have agreed that a "great engineering manager" promotes fairness in their interactions with individual engineers [72]. A key aspect of this fairness is consistently demonstrating appreciation for engineers' contributions.

Software professionals, especially managers and leaders, should understand the factors that drive voice and silence, as communication may be best determined by explicitly considering both. However, further research is needed to understand how different development environments shape voice and silence behaviors.

6.2. Research Implications

This study highlights the importance of research design, particularly the level of analysis. Voice and silence may appear as opposites in studies that frame them as a binary choice on a single issue or when moderating factors influence their relationship [7]. In contrast, our study distinguishes between promotive and prohibitive voice and general, non-motive-based silence. Although it is expected that different operationalizations and methodological factors result in a pattern of correlations in line with our overall findings, the magnitude or strength of the distinctions between voice and silence may vary.

It is important to consider how to increase precision beyond using previously validated scales. For instance, measures for our study should avoid items that assume voice and silence are opposites. Moreover, scale formats should align with the inherent characteristics of the focal constructs being measured. In this sense, frequency scales (e.g., never–frequently) may be appropriate for some constructs, like behaviors, while they may be less appropriate for other constructs like attitudes [73]. Voice and silence have been measured using both frequency and agreement scales (e.g., strongly disagree–strongly agree). Consequently, we use frequency

scales to measure voice and silence since they produce more consistent effects compared to agreement anchors, according to Sherf et al. [10].

Finally, measuring and reporting findings for both voice and silence in SE studies, regardless of the primary research focus, can provide deeper insights into their converging and diverging effects. This approach can foster a more comprehensive understanding and support the balanced advancement of research in this area.

6.3. Future Work

Our study shows that BIS/BAS provides a useful lens for understanding the potential relationship among social-relational concepts. Consequently, further research should examine how other human factors related to software development, such as power dynamics, organizational justice, and team maturity may activate these systems and whether they lead to systematic differences in voice and silence. In this vein, a previous study conducted by Morrison et al. [74] argues that elevated power activates processes associated with BAS whereas a lack of power activates processes associated with BIS. Their findings suggest that, although silence is rooted in the psychological experience of powerlessness, perceived openness of the target can mitigate this effect, encouraging individuals to speak up even when they might otherwise remain silent. However, this study was not conducted within a software engineering context.

In our study, we propose an approach to induce regularization and prevent overfitting by including validity and internal consistency requirements as constraints in the process of optimizing a model (see section 5.4). This approach is motivated by the fact that nontrivial construct validity issues require an iterative process to achieve model optimization. However, further research is needed to provide empirical evidence of the goodness of fit indices across different sample sizes.

Finally, future research should continue to explore how different contexts shape voice and silence. For example, it would be interesting to explore their role in risk assessment within the cybersecurity workforce [75] or in the development of cyber–physical systems that interact with and control the physical world [76]. Additionally, exploring communication strategies through the lens of Hirschman's Exit-Voice-Loyalty framework [77] could yield new insights. Comparing the SE context with other professional domains may also reveal factors that amplify or mitigate the effects of voice and silence. In particular, future work based on qualitative approaches could shed light on the distinctions between voice and silence and their relevance within the SE practices and activities.

6.4. Limitations and Threats to Validity

This study should be interpreted with awareness of its limitations. The threats to validity are organized according to the four types proposed by Wohlin et al. [78].

Internal validity. As the concepts under study involve complex interactions of socio-technical factors, we use multivariate techniques such as EFA, CFA, and SEM to validate our findings. However, a potential limitation of our study lies in its ability to infer causality. While SEM can demonstrate correlations among variables, establishing causality also requires evidence of precedence and the exclusion of alternative (third-variable) explanations. Two aspects of our study mitigate this limitation. First, our hypotheses incorporate causal assumptions [10] as part of the model, which provides a basis for causal assertions [79]. Second, other variable explanations cannot be discounted, e.g., individual differences like personality traits or cultural context. Extroverted professionals may voice frequently and display little silence, resulting in stronger negative correlations between voice and silence. They are also more likely to perceive more impact and worry less about speaking up. To strengthen internal validity, future research should consider utilizing randomized experimental or at least observational longitudinal designs that better satisfy the empirical criteria for causal inference, such as establishing temporal precedence, controlling confounders, and ruling out alternative explanations. Given the limitations of our cross-sectional design, we interpret the estimated path coefficients as associations framed within a theoretical causal model, rather than as evidence of definitive causal effects.

Conclusion validity. We conducted preliminary tests to determine the suitability of performing EFA which helped us identify latent variables. These factors were then confirmed through CFA, and model validity was assessed by using multiple fit indices to ensure alignment with the data before interpreting results. These indices confirmed the overall soundness of the model. SEM enhances conclusion validity by testing the model as a whole, rather than evaluating individual hypotheses. Although SEM also corrects measurement error by inferring latent variables from observable variables, the main remaining threat to conclusion validity is overfitting the structural model. To address this from a statistical standpoint, we conducted a bootstrap-based validation using 1000 random 2/3 vs. 1/3 splits to obtain CIs. In each iteration, the SEM from m_3 was refitted independently, and the key fit metrics were extracted. Overall, the model's full-sample fit is fully consistent with the distribution of resampled fits. This supports the stability and generalizability of the structure of our SEM in m_3 and confirms that the structure learned from constrained optimization from Equation (2) did not lead to severe overfitting, as expected due to the introduced constraints.

External Validity. This study prioritized internal validity due to voice and silence are universal aspects of human communication and experience, and their relationships are observable in any sample to some extent. We used convenience and snowball sampling supported by local collaborators to enhance diversity, but introduced potential selection bias, and the sample is not representative of all software professionals. Additionally, restricting the survey to native Spanish-speaking countries limits broader generalization. Consequently, future studies should pursue more balanced, regionally representative samples to strengthen external validity. We made a replication package publicly available in order to enable other researchers to replicate, reproduce and extend our study.

Construct validity. As all six variables are internal experiences that are best reported by individuals themselves [10], self-reported measures were used to capture each construct. To mitigate bias, the survey was voluntary and anonymous, with a one-month recall period to improve response accuracy. To enhance construct validity, we also used validated scales. However, this study posed methodological issues. As noted by Sherf et al. [10], there are different types of silence and different types of voices that can, consequently, lead to different results. In this study, we focus on

After a validity analysis of a sample of 198 software practitioners, we found that the context in which the instrument is administered can impact its validity. Even if the instrument itself is valid, response biases can introduce errors in the measurement. Therefore, a second pilot study was conducted after redefining the sampling strategy. This study involved 158 software practitioners, and the construct validity improved. However, the design faces mono-method bias, common method bias and potential model misspecification due to reliance on a single survey instrument, subjective one-time responses, shared measurement context, and assumptions embedded in the specified factor structure, particularly those related to linearity and independence (among certain constructs) as optimized through Equation (2). While we recognize these limitations and urge cautious interpretation of our findings, they do not undermine the core modeling contribution of this study. Further investigation of these limitations is encouraged in future studies.

7. Conclusion

Our study's results indicate that voice and silence are distinct constructs, even though they may seem to describe the same behavior in different ways, i.e, voice as the expression of ideas and concerns, and silence as their suppression. Although our sample was small for testing our hypotheses, the fact that we still found support for our proposed relationships suggests that our study provides a conservative

demonstration of underlying relationships that may be even more robust in other contexts.

This study calls for more research attention to voice and silence as an important communication behavior in software development. Therefore, we hope this study inspires more research on why, when, and how voice and silence affect software professionals across teams, projects and organization levels. Understanding these differences, organizations could employ different interventions for voice and silence to lower software professional withdrawal and increase performance.

Compliance with ethical standards

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.

CRediT authorship contribution statement

Mary Sánchez-Gordón: Conceptualization, Methodology, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing; Visualization, Resources. Ricardo Colomo-Palacios: Conceptualization, Methodology, Validation, Writing – review & editing. Alex Sanchez Gordon: Investigation, Writing – review & editing. Aliaksandr Hubin: Formal analysis, Writing – original draft, Writing – review & editing: Visualization.

Declaration of competing interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgments

We thank the software professionals who participated in this study. We also thank Cristian Timbi and Juan Carlos Perez Arriaga for their collaboration in the dissemination of the survey.

Data Availability

A comprehensive replication package including our (anonymous) dataset, instruments and analysis scripts is stored in an open data archive at Figshare [11].

References

[1] P. Lenberg, R. Feldt, L.G. Wallgren, Behavioral software engineering: A definition and systematic literature review, Journal of Systems and Software 107 (2015) 15–37. https://doi.org/10.1016/j.jss.2015.04.084.

[2] J. Menezes, C. Gusmão, H. Moura, Risk factors in software development projects: a systematic literature review, Software Qual J 27 (2019) 1149–1174. https://doi.org/10.1007/s11219-018-9427-5.

[3] R. Vallon, B.J. da Silva Estácio, R. Prikladnicki, T. Grechenig, Systematic literature review on agile practices in global software development, Information and Software Technology 96 (2018) 161–180. https://doi.org/10.1016/j.infsof.2017.12.004.

[4] C. Matthies, M. Sánchez-Gordón, J.B. Jørgensen, L. Prechelt, "Communication Is a Scarce Resource!'': A Summary of CHASE'22 Conference Discussions, (2022). https://doi.org/10.48550/arXiv.2207.00054.

[5] T. Mens, M. Cataldo, D. Damian, The Social Developer: The Future of Software Development [Guest Editors' Introduction], IEEE Software 36 (2019) 11–14. https://doi.org/10.1109/MS.2018.2874316.

[6] M.O. Ahmad, T. Gustavsson, A. Katin, N. Taušan, V. Mandić, Non- Technical Aspects of Technical Debt in the Context of Large-Scale Agile Development: A Qualitative Study, in: 2024 50th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2024: pp. 260–267. https://doi.org/10.1109/SEAA64295.2024.00048.

[7] E.W. Morrison, Employee Voice and Silence: Taking Stock a Decade Later, Annual Review of Organizational Psychology and Organizational Behavior 10 (2023) 79–107. https://doi.org/10.1146/annurev-orgpsych-120920-054654.

[8] M. Zhan, When to speak up at work: a review of employee voice and silence behavior using a prospect approach, Atlantic Journal of Communication 28 (2020) 273–288. https://doi.org/10.1080/15456870.2020.1720682.

[9] T. Lorey, P. Ralph, M. Felderer, Social science theories in software engineering research, in: Proceedings of the 44th International Conference on Software

Engineering, Association for Computing Machinery, New York, NY, USA, 2022: pp. 1994–2005. https://doi.org/10.1145/3510003.3510076.

[10] E.N. Sherf, M.R. Parke, S. Isaakyan, Distinguishing Voice and Silence at Work: Unique Relationships with Perceived Impact, Psychological Safety, and Burnout, AMJ 64 (2021) 114–148. https://doi.org/10.5465/amj.2018.1428.

[11] M. Sánchez-Gordón, R. Colomo-Palacios, A. Sanchez Gordon, A. Hubin, Online appendix: The Role of Voice and Silence Behaviors in Software Development: A Structural Equation Modeling Analysis, (2025). https://figshare.com/s/e7a548858bbaaaddaa72.

[12] A. Wilkinson, J. Donaghey, T. Dundon, R.B. Freeman, Handbook of Research on Employee Voice: Elgar original reference, Edward Elgar Publishing, 2014.

[13] E. Morrison, Employee Voice and Silence, Annual Review of Organizational Psychology and Organizational Behavior 1 (2014) 173–197. https://doi.org/10.1146/annurev-orgpsych-031413-091328.

[14] A.N. Li, S. Tangirala, How employees' voice helps teams remain resilient in the face of exogenous change, Journal of Applied Psychology (2022). https://doi.org/10.1037/apl0000874.

[15] J.R. Detert, A.C. Edmondson, Implicit Voice Theories: Taken-for-Granted Rules of Self-Censorship at Work, AMJ 54 (2011) 461–488. https://doi.org/10.5465/amj.2011.61967925.

[16] P. Holland, A. Pyman, B.K. Cooper, J. Teicher, Employee voice and job satisfaction in Australia: The centrality of direct voice, Human Resource Management 50 (2011) 95–111. https://doi.org/10.1002/hrm.20406.

[17] C. Rees, K. Alfes, M. Gatenby, Employee voice and engagement: connections and consequences, The International Journal of Human Resource Management 24 (2013) 2780–2798. https://doi.org/10.1080/09585192.2013.763843.

[18] A. Newman, R. Donohue, N. Eva, Psychological safety: A systematic review of the literature, Human Resource Management Review 27 (2017) 521–535. https://doi.org/10.1016/j.hrmr.2017.01.001.

[19] J. Donaghey, N. Cullinane, T. Dundon, A. Wilkinson, Reconceptualising employee silence: problems and prognosis, Work, Employment and Society 25 (2011) 51–67. https://doi.org/10.1177/0950017010389239.

[20] E. Nechanska, E. Hughes, T. Dundon, Towards an integration of employee voice and silence, Human Resource Management Review 30 (2020) 100674. https://doi.org/10.1016/j.hrmr.2018.11.002.

[21] S. Petter, If You Can't Say Something Nice: Factors Contributing to Team Member Silence in Distributed Software Project Teams, in: Proceedings of the 2018 ACM SIGMIS Conference on Computers and People Research, ACM, Buffalo-Niagara Falls NY USA, 2018: pp. 43–49. https://doi.org/10.1145/3209626.3209718.

[22] T.L. Roberts, P.B. Lowry, P.D. Sweeney, An evaluation of the impact of social presence through Group size and the use of collaborative software on Group member "Voice" in face-to-face and computer-mediated task Groups, IEEE Transactions on Professional Communication 49 (2006) 28–43. https://doi.org/10.1109/TPC.2006.870460.

[23] M. Greiler, M.-A. Storey, A. Noda, An Actionable Framework for Understanding and Improving Developer Experience, IEEE Transactions on Software Engineering (2022) 1–1. https://doi.org/10.1109/TSE.2022.3175660.

[24] T.R. Tulili, A. Capiluppi, A. Rastogi, Burnout in software engineering: A systematic mapping study, Information and Software Technology 155 (2023) 107116. https://doi.org/10.1016/j.infsof.2022.107116.

[25] L. Hunt, M.A. Ferrario, A review of how whistleblowing is studied in software engineering, and the implications for research and practice, in: Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society, Association for Computing Machinery, New York, NY, USA, 2022: pp. 12–23. https://doi.org/10.1145/3510458.3513013.

[26] M. Knoll, J. Wegge, C. Unterrainer, S. Silva, T. Jønsson, Is our knowledge of voice and silence in organizations growing? Building bridges and (re)discovering opportunities, German Journal of Human Resource Management 30 (2016) 161–194. https://doi.org/10.1177/2397002216649857.

[27] C.S. Carver, Approach, Avoidance, and the Self-Regulation of Affect and Action, Motiv Emot 30 (2006) 105–110. https://doi.org/10.1007/s11031-006-9044-7.

[28] G.M. Spreitzer, Psychological Empowerment in the Workplace: Dimensions, Measurement, and Validation, The Academy of Management Journal 38 (1995) 1442–1465.

[29] E.H. Schein, W.G. Bennis, Personal and Organizational Change Through Group Methods: The Laboratory Approac, Wiley, New York, 1965. https://doi.org/10.1177/074171366701700211 (accessed August 23, 2023).

[30] A. Edmondson, Psychological Safety and Learning Behavior in Work Teams, Administrative Science Quarterly 44 (1999) 350–383. https://doi.org/10.2307/2666999.

[31] A.C. Edmondson, Z. Lei, Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct, Annual Review of Organizational Psychology and Organizational Behavior 1 (2014) 23–43. https://doi.org/10.1146/annurev-orgpsych-031413-091305.

[32] M.L. Frazier, S. Fainshmidt, R.L. Klinger, A. Pezeshkan, V. Vracheva, Psychological Safety: A Meta-Analytic Review and Extension: PERSONNEL PSYCHOLOGY, Personnel Psychology 70 (2017) 113–165. https://doi.org/10.1111/peps.12183.

[33] A. Tkalich, D. Šmite, N.H. Andersen, N.B. Moe, What Happens to Psychological Safety When Going Remote?, IEEE Software (2022) 1–9. https://doi.org/10.1109/MS.2022.3225579.

[34] A. Alami, M. Zahedi, O. Krancher, Antecedents of psychological safety in agile software development teams, Information and Software Technology 162 (2023) 107267. https://doi.org/10.1016/j.infsof.2023.107267.

[35] A. Alami, M. Zahedi, O. Krancher, The role of psychological safety in promoting software quality in agile teams, Empir Software Eng 29 (2024) 119. https://doi.org/10.1007/s10664-024-10512-1.

[36] M. Greiler, M.-A. Storey, A. Noda, An Actionable Framework for Understanding and Improving Developer Experience, IEEE Transactions on Software Engineering 49 (2023) 1411–1425. https://doi.org/10.1109/TSE.2022.3175660.

[37] F. Fagerholm, M. Ikonen, P. Kettunen, J. Münch, V. Roto, P. Abrahamsson, Performance Alignment Work: How software developers experience the continuous adaptation of team performance in Lean and Agile environments, Information and Software Technology 64 (2015) 132–147. https://doi.org/10.1016/j.infsof.2015.01.010.

[38] P. Lenberg, R. Feldt, Psychological safety and norm clarity in software engineering teams, in: Proceedings of the 11th International Workshop on Cooperative and Human Aspects of Software Engineering, Association for Computing Machinery, New York, NY, USA, 2018: pp. 79–86. https://doi.org/10.1145/3195836.3195847.

[39] T. Dingsøyr, T. Dybå, Team effectiveness in software development: Human and cooperative aspects in team effectiveness models and priorities for future studies, in: 2012 5th International Workshop on Co-Operative and Human Aspects of Software Engineering (CHASE), 2012: pp. 27–29. https://doi.org/10.1109/CHASE.2012.6223016.

[40] D. Strode, T. Dingsøyr, Y. Lindsjorn, A teamwork effectiveness model for agile software development, Empirical Softw. Engg. 27 (2022). https://doi.org/10.1007/s10664-021-10115-0.

[41] L. Koopmans, Measuring Individual Work Performance, Doctoral thesis, TNO Innovation for Life – VU University Medical Center, 2015.

[42] L. Koopmans, C. Bernaards, V. Hildebrandt, S. Buuren, A. van der Beek, H. De Vet, Improving the Individual Work Performance Questionnaire using Rasch Analysis., Journal of Applied Measurement 15 (2014) 160–75. https://doi.org/10.1136/oemed-2013-101717.51.

[43] W.B. Schaufeli, T.W. Taris, The conceptualization and measurement of burnout: Common ground and worlds apart, Work & Stress 19 (2005) 256–262. https://doi.org/10.1080/02678370500385913.

[44] L.R. Bolton, R.D. Harvey, M.J. Grawitch, L.K. Barber, Counterproductive Work Behaviours in Response to Emotional Exhaustion: A Moderated Mediational Approach, Stress and Health 28 (2012) 222–233. https://doi.org/10.1002/smi.1425.

[45] A.G. Miner, T.M. Glomb, State mood, task performance, and behavior at work: A within-persons approach, Organizational Behavior and Human Decision Processes 112 (2010) 43–57. https://doi.org/10.1016/j.obhdp.2009.11.009.

[46] D. Graziotin, F. Fagerholm, X. Wang, P. Abrahamsson, Consequences of Unhappiness while Developing Software, in: 2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion), 2017: pp. 42–47. https://doi.org/10.1109/SEmotion.2017.5.

[47] D.A. Shepherd, H. Patzelt, C.M. Berry, Why Didn't You Tell Me? Voicing Concerns Over Objective Information About a Project's Flaws, Journal of Management 45 (2019) 1087–1113. https://doi.org/10.1177/0149206316688942.

[48] S. Tangirala, R. Ramanujam, Ask and You Shall Hear (but Not Always): Examining the Relationship Between Manager Consultation and Employee Voice, Personnel Psychology 65 (2012) 251–282. https://doi.org/10.1111/j.1744-6570.2012.01248.x.

[49] J. Liang, C.I.C. Farh, J.-L. Farh, Psychological Antecedents of Promotive and Prohibitive Voice: A Two-Wave Examination, AMJ 55 (2012) 71–92. https://doi.org/10.5465/amj.2010.0176.

[50] R. O'Donovan, E. McAuliffe, A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour, BMC Health Services Research 20 (2020) 101. https://doi.org/10.1186/s12913-020-4931-2.

[51] T.E. Becker, G. Atinc, J.A. Breaugh, K.D. Carlson, J.R. Edwards, P.E. Spector, Statistical control in correlational studies: 10 essential recommendations for organizational researchers, Journal of Organizational Behavior 37 (2016) 157–167. https://doi.org/10.1002/job.2053.

[52] C. Kiewitz, S.L.D. Restubog, M.K. Shoss, P.R.J.M. Garcia, R.L. Tang, Suffering in silence: Investigating the role of fear in the relationship between abusive supervision and defensive silence., Journal of Applied Psychology 101 (2016) 731–742. https://doi.org/10.1037/apl0000074.

[53] M. Knoll, R. Hall, O. Weigelt, A Longitudinal Study of the Relationships Between Four Differentially Motivated Forms of Employee Silence and Burnout, Journal of Occupational Health Psychology 24 (2018). https://doi.org/10.1037/ocp0000143.

[54] R. Shaukat, A. Khurshid, Woes of silence: the role of burnout as a mediator between silence and employee outcomes, Personnel Review 51 (2021) 1570–1586. https://doi.org/10.1108/PR-07-2020-0550.

[55] L.F. Röllmann, M. Weiss, H. Zacher, Does Voice Benefit or Harm Occupational Well-Being? The Role of Job Insecurity, British Journal of Management 32 (2021) 708–724. https://doi.org/10.1111/1467-8551.12471.

[56] R.E. Johnson, C.-H. Chang, T. Meyer, K. Lanaj, J. Way, Approaching Success or Avoiding Failure? Approach and Avoidance Motives in the Work Domain, European Journal of Personality 27 (2013) 424–441. https://doi.org/10.1002/per.1883.

[57] M. Chamberlin, D.W. Newton, J.A. LePine, A meta-analysis of empowerment and voice as transmitters of high-performance managerial practices to job performance, Journal of Organizational Behavior 39 (2018) 1296–1313. https://doi.org/10.1002/job.2295.

[58] W.E.K. Lehman, D.D. Simpson, Employee substance use and on-the-job behaviors. - PsycNET, Journal of Applied Psychology 77 (1992) 309–321.

[59] P.J. Ramos-Villagrasa, J.R. Barrada, E. Río, L. Koopmans, Assessing Job Performance Using Brief Self-report Scales: The Case of the Individual Work Performance Questionnaire, Revista de Psicología Del Trabajo y de Las Organizaciones 35 (2019) 195–205. https://doi.org/10.5093/jwop2019a21.

[60] L. Hu, P.M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives, Structural Equation Modeling: A Multidisciplinary Journal 6 (1999) 1–55. https://doi.org/10.1080/10705519909540118.

[61] A. Hubin, G. Storvik, Mode jumping MCMC for Bayesian variable selection in GLMM, Computational Statistics & Data Analysis 127 (2018) 281–297. https://doi.org/10.1016/j.csda.2018.05.020.

[62] S. Faraj, A. Yan, Boundary work in knowledge teams, Journal of Applied Psychology 94 (2009) 604–617. https://doi.org/10.1037/a0014367.

[63] S. Kim, H. Lee, T.P. Connerton, How Psychological Safety Affects Team Performance: Mediating Role of Efficacy and Learning Behavior, Front Psychol 11 (2020) 1581. https://doi.org/10.3389/fpsyg.2020.01581.

[64] L. Barros, C. Tam, J. Varajão, Agile software development projects–Unveiling the human-related critical success factors, Information and Software Technology 170 (2024) 107432. https://doi.org/10.1016/j.infsof.2024.107432.

[65] S. Maji, S. Dixit, Exploring Self-Silencing in Workplace Relationships: A Qualitative Study of Female Software Engineers, The Qualitative Report 25 (2020) 1505–1525.

[66] S. Thorgren, E. Caiman, The Role of Psychological Safety in Implementing Agile Methods across Cultures, Research-Technology Management 62 (2019) 31–39. https://doi.org/10.1080/08956308.2019.1563436.

[67] B. Santana, S. Freire, J.A.M. Santos, M. Mendonça, Psychological Safety in the Software Work Environment, IEEE Software 41 (2024) 86–94. https://doi.org/10.1109/MS.2024.3386532.

[68] M.O. Ahmad, Psychological Safety, Leadership and Non-Technical Debt in Large-Scale Agile Software Development, in: 2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS), 2023: pp. 327–334. https://doi.org/10.15439/2023F8595.

[69] Z. Feng, I. Steinmacher, M. Gerosa, T. Menezes, A. Serebrenik, R. Milewicz, A. Sarma, The Multifaceted Nature of Mentoring in OSS: Strategies, Qualities, and Ideal Outcomes, in: 2025 IEEE/ACM 18th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE), 2025: pp. 203–214. https://doi.org/10.1109/CHASE66643.2025.00031.

[70] K. Silva, A. Barcomb, R.D.S. Santos, Paths to Testing: Why Women Enter and Remain in Software Testing?, in: Companion Proceedings of the 32nd ACM International Conference on the Foundations of Software Engineering, ACM, Porto de Galinhas Brazil, 2024: pp. 4–9. https://doi.org/10.1145/3663529.3663822.

[71] E. Sesari, F. Sarro, A. Rastogi, It is Giving Major Satisfaction: Why Fairness Matters for Developers, It Is Giving Major Satisfaction (2024).

[72] E. Kalliamvakou, C. Bird, T. Zimmermann, A. Begel, R. DeLine, D.M. German, What Makes a Great Manager of Software Engineers?, IEEE Transactions on Software Engineering (2018) 1–1. https://doi.org/10.1109/TSE.2017.2768368.

[73] P.M. Podsakoff, N.P. Podsakoff, L.J. Williams, C. Huang, J. Yang, Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix, Annual Review of Organizational Psychology and Organizational Behavior 11 (2024) 17–61. https://doi.org/10.1146/annurev-orgpsych-110721-040030.

[74] E.W. Morrison, K.E. See, C. Pan, An Approach-Inhibition Model of Employee Silence: The Joint Effects of Personal Sense of Power and Target Openness, Personnel Psychology 68 (2015) 547–580. https://doi.org/10.1111/peps.12087.

[75] A. Brilingaitė, L. Bukauskas, I. Domarkienė, T. Rančelis, L. Ambrozaitytė, R. Pirta, R.G. Lugo, B.J. Knox, Towards projection of the individualised risk assessment for the cybersecurity workforce, Computer Standards & Interfaces 93 (2025) 103962. https://doi.org/10.1016/j.csi.2024.103962.

[76] J.M. Fernandes, J.E. Rivadeneira, A. Rodrigues, F. Boavida, J. Sá Silva, People 4.0 — A model for Human-in-the-Loop CPS-based systems, Computer Standards & Interfaces 91 (2025) 103895. https://doi.org/10.1016/j.csi.2024.103895.

[77] A.O. Hirschman, Exit, Voice, and Loyalty, Harvard University Press, Cambridge, Mass, 1970.

[78] C. Wohlin, P. Runeson, M. Host, M.C. Ohlsson, B. Regnell, A. Wesslen, Experimentation in Software Engineering, Springer Science & Business Media, 2012.

[79] K.A. Bollen, J. Pearl, Eight Myths About Causality and Structural Equation Models, in: S.L. Morgan (Ed.), Handbook of Causal Analysis for Social Research, Springer Netherlands, Dordrecht, 2013: pp. 301–328. https://doi.org/10.1007/978-94-007-6094-3_15.