The role of gender and psychological safety in effective Agile teams:
An exploratory analysis

Author: Evelien Heijdens
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

Organisations are continuously looking for ways to improve and be successful, and a way to achieve this is through the Agile way of working. Psychological safety has also shown to be of great importance to a team's success and research has noted that men and women perceived psychological safety differently. However, studies are limited when it comes to studying the combination of psychological safety and gender, even more when psychological safety is innovatively investigated through the exploratory manner of video observations. Therefore, by analyzing 4 retrospective meetings of effective Agile teams with 21 males and 6 females, this research provides an extensive exploration of 8 psychological safety behaviours: *voice behaviours*, *defensive voice behaviours*, *silence behaviours*, *defensive silence behaviours*, *collaboration behaviours*, *unsupportive behaviours*, *learning or improvement behaviours and familiarity behaviours*. Whilst, quantitative comparative analysis revealed no significant difference between men and women with regards to all psychological safety behaviours, but *familiarity behaviours*, frequency counts showed differences between the number of expressed behaviours by men and women, so that, overall, women display more *voice behaviours*, *learning or improvement behaviours and familiarity behaviours* as well as fewer *silence behaviours*, *defensive silence behaviours* and *unsupportive behaviours* compared to men. Theoretically, this research contributes to a better understanding of the influence of gender on psychological safety in a team. Practically, this research shows how women in a team, even when a minority, can still influence team dynamics and thus could help managers when composing teams.

Graduation Committee members:
Dr. L. Carminati
Dr. D. H. van Dun

Keywords
Agile, Gender, Psychological safety, Team Effectiveness, Video Observations

1. INTRODUCTION

Creating a safe and pleasant environment where employees feel fulfilled and satisfied is important for organizations since employees who feel happy and satisfied can perform better (Abrudan & Conea-Simiuc, 2019). One way to achieve better performance is by working Agile (Melnik & Maurer, 2006; Begel & Nagappan, 2007). Even though the Agile way of working finds its origin in the software development industry (Fowler & Highsmith, 2001), it has increased in popularity and been adopted in other fields thanks to its focus on team effectiveness (Grzeszczyk, 2020; Hoeseb & Tanner, 2020; Scott et al., 2021).

The Agile way of working is unique and indeed is centered around teams. Agile teams are self-organizing teams that are composed of "individuals [that] manage their own workload, shift work among themselves based on need and best fit and participate in team decision making" (Hoda et al., 2012, p. 426). In these teams there is no entitled leader, and leadership emerges rather in a natural manner among the team members (Przybilla et al., 2019). Agile teams have become increasingly popular in today's organisations due to their focus on boosting team effectiveness. Besides autonomy and flexibility, research has investigated several other factors that may impact Agile teams' effectiveness, such as high task interdependence, cross-fertilization, self-transcendence and performance effectiveness (Acharya & Colomo-Palacios, 2019). Although explored little within the Agile environment, other factors may also influence team effectiveness, such as gender diversity and psychological safety (Lee et al., 2018).

The gender composition of a team has the potential to affect whether working conditions are perceived as unpleasant or as uplifting (Ivanova-Stenzel & Kübler, 2011). Since the male and female brain function differently (Zaidi, 2010), men and women tend to show differences in their behaviour, especially given that decision-making in team formation has become the status quo over the years (Seong & Hong, 2014). For instance, Lee et al.'s (2018) research has shown that the way conflicts are managed is different depending on the gender composition of a team. Research has also found that gender composition and psychological safety can influence employee satisfaction, which then leads to a better team performance (Apesteguia et al., 2012). According to Woolley et al. (2010), teams with a greater percentage of women perform better since the average level of social sensitivity of the group members is higher. Similarly, Rogelberg and Rumery (1996) found that lone female teams outperformed all-male teams. Other research has noted that business teams with an equal gender mix perform better than male-dominated teams, especially when it comes to sales and profits (Hoogendoorn et al., 2013), and underlined that gender-mixed teams stimulate novel solutions and radical innovation (Díaz-García et al., 2013).

The abovementioned insights of researchers show that gender diversity plays an important role and is therefore a factor that allows teams to flourish. Similarly, psychological safety seems to impact firm performance (Edmondson & Lei, 2014). According to Edmondson (2002), psychological safety means that no one will be punished or humiliated for errors, questions, or requests for help, in the service of reaching ambitious performance goals. In other words, the team is safe for interpersonal risk taking. Psychological safety has recently been divided in different subdimensions, namely *voice behaviours, defensive voice behaviours, silence behaviours, defensive silence behaviours, supportive/collaboration behaviours, unsupportive behaviours, learning or improvement-oriented behaviours and familiarity behaviours* (O'Donovan et al., 2020).

Traditionally, psychological safety has been studied as a perceived variable, but more research has been calling for psychological safety to be studied with a wider, holistic approach (O'Donovan et al., 2020). Indeed, nowadays research is inclined to more objective ways of research, rather than relying solely on surveys, and spur to use more innovative methods, such as video observations (Zhao et al, 2019). Yet, research implementing these alternative methods are still scarce. In this thesis, team members' verbal as well as nonverbal behaviours are explored thoroughly by analysing video observations (O'Donovan et al, 2020). Hence, since how women differ from men concerning *observed* psychological safety behaviours has not been researched yet, especially in the environment of agile teams in which women are a minority (Warnert, 2015), the following research question is formulated:

"How may observed psychological safety vary between gender in effective Agile teams?"

The overall aim of this thesis is thus to explore how psychological safety behaviours may vary between males and females in Agile teams. In order to reach this aim, the newest version of the Psychological Safety codebook (O'Donovan et al., 2020) is used to code the retrospective meetings of four Agile teams. This codebook identifies behaviours related to psychological safety which can be innovatively observed and measured.

By exploring the relationship between observed psychological safety and gender in Agile teams, this thesis contributes to the Agile and psychological safety literature in two ways: Firstly, by studying psychological safety through an innovative and needed methodology, i.e., video observations (Zhao et al, 2019), we address the paucity of research calling for more objective measurements; Secondly, the combination of observed psychological safety and gender has scarcely been explored before in Agile contexts. Since Agile teams are gaining in popularity (Dikert et al., 2016), it is crucial to research how psychological safety and gender play a role in these types of teams since this can contribute to a better understanding of their influence on a team's effectiveness and success. As for practical implications, the results of this thesis can help managers to better understand and take into consideration the importance of gender diversity and psychological safety when composing a team in order to maximize the team's potential.

The remaining part of this thesis is structured as follows. Firstly, the theoretical framework and methodology are presented in the upcoming chapters. Secondly, the results are reported, followed by a discussion of the gained knowledge. The limitations and future research of the research are then discussed and overall conclusion.

2. THEORETICAL FRAMEWORK

The literature explored for this research covers the Agile way of working, followed by an in-depth analysis of the role of gender in teams in general and then continuing with the specific importance of females in teams. Lastly, the different behaviours showed during the recorded meetings are defined and explained.

2.1 Agile & Team Effectiveness

The Agile way of working is a philosophy and a methodology that has its origin in the software development field. In 2001 the Manifesto for Agile Software Development was published by software professionals (Fowler & Highsmith, 2001). Individuals and interactions, consumer collaboration, functional software, and, lastly, responding to change are the four primary characteristics addressed in this manifesto (Fowler & Highsmith,

2001). Due to the proven success of this method (Dikert et al. 2016), organizations in other fields have started to implement it, from service industries to HR practices, including the financial sector (Grzeszczyk, 2020; Hoeseb & Tanner, 2020; Scott et al., 2021). The main reason for these organisations to implement the Agile way of working is its positive impact on group effectiveness (Acharya & Colomo-Palacios, 2019) which stems, mainly, from the cross-disciplinary and self-managed teams. These work in sprints, consisting of three meetings; a planning, refinement and retrospective meeting. Agile teams are indeed characterized by high level of autonomy, regularly reflecting and adapting their behaviour in line with the present circumstances (Fowler & Highsmith, 2001). This research also provides a reflection of effective Agile teams. It reflects on the team composition regarding gender and its possible influence on a team.

2.2 The influence of Gender in (Agile) Teams

Males and females are wired differently (Elliot, 2011). This statement has been researched thoroughly for decades and translates into the working environment in which males and females behave differently. One theory that explains the difference between how a male and female brain functions is the empathizing-systemizing theory. This theory claims that the female brain is primarily hard-wired for empathy, while the male brain is primarily hard-wired for understanding and building systems (Baron-Cohen, 2005). Along a similar line, Zaidi (2010) speaks of differences in the way males and females learn already during their educational careers: females seem to prefer using the brain areas responsible for verbal and emotional functioning, whereas males have a greater tendency to use areas of the brain that are concerned with spatial and mechanical tasks. This is why males tend to "space out" during a lesson (Zaidi, 2010).

When translated into the work context, such gender differences can likewise influence team dynamics. Perhaps given the original and subsequent areas of development of the Agile way of working, which are usually associated with men (Halliday et al., 2022), research has noted that women in Agile teams tend to be a minority (Warnert, 2015). Similarly, women are a minority in the financial sector (Fair Bank Guide, 2020). Yet, studies have highlighted the importance of women in teams as their presence improves team collaboration (Ortmann et al., 1999; Bear & Woolley, 2011) as well as their peers' participation and involvement (Eagly & Johnson, 1990). This is of particular importance in self-managed teams since the absence of a formal leader makes everyone peer and gives each team member the possibility to contribute equally to the team's project and effectiveness (Przybilla et al., 2019). Since research on females in Agile teams is rather scarce, we need to rely on research on teams to better understand their role and impact on a team.

Research shows that trust and reciprocity is higher among male-female pairings, especially when males are the sender and women the receiver, compared to same sex pairings, which again highlights the importance of the presence of women in a team (Chaudhuri, 2003). In addition, women tend to show lower levels of trust which might be linked to a higher level of risk aversion (Chaudhuri, 2003). Andreoni and Vesterlund's (2001) analysis confirms these findings as they found that women behave differently in bargaining games as they are more sensitive to the economic costs of being generous in negotiations (Andreoni & Vesterlund, 2001). Additionally, females adopt problem-focused coping methods such as planning, communication, and technique-oriented coping more often than males (Nicholls et al., 2007). In conclusion, the fact that men and women are wired differently already shows its first effects during the educational career, after which it continues during their working career,

including in a team formation with various research showing that it is beneficial to have women on your team since they enhance the team dynamics.

2.3 Differences in Male and Female Psychological Safety Behaviours

Since males and females differ and favor different behaviors, this may have an impact on how they express themselves within a team which, ultimately, may affect the team's psychological safety. Edmondson defines psychological safety as a shared belief that the team is safe for interpersonal risk-taking (Edmondson, 2002). Even though research exploring gender and psychological safety directly is rather scarce, some studies do link the two concepts highlighting that women seem to have a positive influence on their team members and therefore on the team as a whole (Lee et al., 2018; Chaudhuri 2002; Nicholls et al., 2007; Ortmann et al., 1999; Bear & Woolley, 2011; Briton & Hall, 1995; Phutela, 2015). This will be further elaborated on in the next subchapters.

However, these studies have generally explored *perceived* psychological safety through survey methods rather than more objective ways to capture this construct, like video observations (O'Donovan et al., 2020; Zhao et al, 2019). This thesis addresses this need by considering *observed* psychological safety. Hence, following O'Donovan et al. (2020), psychological safety is divided into the subsequent dimensions: *voice behaviours, defensive voice behaviours, silence behaviours, defensive silence behaviours, collaboration behaviours, unsupportive behaviours, learning or improvement behaviours and familiarity behaviours. Voice behaviours, learning or improvement behaviours, collaboration behaviours and familiarity behaviours* can be seen as positive psychological safety behaviours as they contribute to a higher psychological safety score. Oppositely, *defensive voice behaviours, silence behaviours, defensive silence behaviours and unsupportive behaviours,* can be seen as negative psychological safety behaviours as they contribute to a lower psychological safety score. Below, these subdimensions are explained in detail in relation to gender differences, including a conclusive hypothesis per subdimension. However, in order to explore whether females show more psychological safety overall, as well as show more positive psychological safety, but less negative psychological safety, as implied by research on *perceived* psychological safety, the following hypothesis were advanced:

Hypothesis 1: *"Females show more psychological safety behaviours in total compared to men"*
Hypothesis 1a: *"Females show more positive psychological safety behaviours compared to men"*
Hypothesis 1b: *"Females show less negative psychological safety behaviours compared to men"*

2.3.1 Voice Behaviours & Defensive Voice Behaviours

Morrison et al. (2011) refer to voice as a discretionary behaviour where a message is transferred from a sender to a recipient with an objective of bringing improvement and positive change. Van Dyne and Le Pine (1998) attach the term voice to a group member speaking up with an innovative suggestion for change to a standard operating procedure in order to enhance the workflow in spite of the chance of upsetting other team members. The chance of possibly upsetting other team members is a risk that needs to be calculated before speaking up with a risky suggestion to challenge the status quo. In research it has been shown that disrupting the status quo can lead to great success (Brown & Anthony, 2011). This is therefore an important thing that team members could do in order for companies to flourish in an everchanging market (Van Dyne and Le Pine, 1998; Lam and Mayer, 2014).

Another subdimension is *defensive voice behaviour*. While *voice behaviours* contribute to a high psychological safety score, *defensive voice behaviours* contribute to a low psychological safety score (O'Donovan et al., 2020). According to Van Dyne et al. (2003), *defensive voice behaviours* indicate behaviours that show motives of self-protection. The authors define defensive voice as "expressing work-related ideas, information or opinions – based on fear – with the goal of protecting the self" (Van Dyne et al., 2003, p. 1372). Overall, individuals in self-managed teams show more *voice behaviours* than individuals in traditional managed teams (Le Pine & Van Dyne, 1998).

However, research does not form one uniform opinion on the role of gender in employee *voice behaviours*. Eibl et al. (2020) propose that in male-dominated environments, women might experience lower self-efficacy beliefs, hesitate to speak up, and limit their ideas and suggestions. This is in line with Le Pine and Van Dyne's (1998) results. The authors noted that, in contrast to the popular belief that women talk more than men (Mehl et al., 2007), men engaged in more voice than women in groups (Le Pine & Van Dyne, 1998). On the contrary, other research claims that there is no relationship between gender and voice (Morrison et al., 2011). In order to see whether females indeed show less (*defensive*) *voice behaviour* compared to the opposite gender, the following hypothesis is put forward:

Hypothesis 2: *"Females show less voice behaviour and defensive voice behaviour compared to men"*

2.3.2 Learning or Improvement Behaviours

Team members who experience higher levels of psychological safety are more inclined to speak up, similarly, they are also more likely to identify and share mistakes with their team (Roussin et al., 2018). Edmondson (1999, p. 353) views *learning or improvement behaviours* as "an ongoing process of reflection and action, characterized by asking questions, seeking feedback, experimenting, reflecting on results, and discussing errors or unexpected outcomes of actions". Women are more affected by self-doubt than men with regards to their contributions to learning and their ability (Roussin et al., 2018). For instance, research has found that women outperform men with regards to essential leadership skills on tests, such as emotional awareness and recognizing interpersonal interactions (Case & Oetema-paul, 2015). And yet, leadership positions were and are mainly occupied by men (Case & Oetama-Paul, 2015; Heilman, 2012). This may be of influence on the women in the team as they may have experienced a man being their boss in their working career more often prior to joining an Agile team.

There might be also another reason for women feeling less psychologically safe to speak up about mistakes. According to research it is expected from women to show less misbehaviour as women showing misbehaviour is seen as more inappropriate compared to a man showing misbehaviour (Heilman, 2012). This may refrain women to acknowledge a mistake or to engage in risky behaviour that could lead to a potential mistake. Not only do men judge female misbehaviour more severely compared to male misbehaviour, but men also tend to penalize women more harshly independently of the severity of the misbehaviour (Bowles & Gelfand, 2010). Since literature implies females are less likely to show *learning or improvement behaviours*, the following hypothesis is formulated:

Hypothesis 3: *"Females show less learning or improvement behaviours compared to men"*

2.3.3 Silence Behaviours & Defensive Silence Behaviours

Non-verbal communication, such as facial expression and body language, are very powerful as it greatly affects the whole communication process (Phutela, 2015). Various scholars emphasize withholding as the fundamental aspect of *silence behaviour* (Pinder & Harlos, 2001; Morrison & Milliken, 2000). According to Morrison and Millikin (2000), the main motivation for employees to remain silent on purpose is because of fear. Following Edmonson's (1999) work on psychological safety, this thus would mean that the employee does not feel psychologically safe to speak up. Van Dyne et al. (2003) covers the definition of *silence behaviour* by saying employees are silent as a result of being fundamentally disengaged or not having the feeling of being able to make a difference in the group. These behaviours include non-verbal behaviour, e.g., closed body language or not making eye contact as well as not speaking up. According to research, not only do women show signs of positive non-verbal behavior, such as nodding, smiling as well as making eye contact more frequently than men (Case & Oetema-Paul, 2015; Phutela, 2015), but they also have a greater ability to interpret non-verbal signals compared to men (Phutela, 2015).

This research, following O'Donovan et al.'s (2020) subdimensions only includes nonverbal, *silence behaviours* (i.e., facial expression or body language that indicates fear, facial expression or body language indicates disengagement or indifference, and sudden closed body language) which can diminish the psychological safety. Furthermore, while in some situations *silence behaviour* can be passive (Van Dyne et al., 2003), *defensive silence* is always proactive behaviour with a purpose of protecting oneself from external threats. Van Dyne et al. (2003) defined *defensive silence behaviours* as: "withholding relevant ideas, information, based on fear that expression of ideas is personally risky." (Van Dyne et al., 2003, p. 1367). Besides showing aggressive body language, *defensive silence behaviours* also include reacting cold to or ignoring a joke (O'Donovan et al., 2020). Research shows that, stereotypically, men tend to show more aggressive behaviour (Phutela, 2015). Based on the research the following hypothesis is advanced:

Hypothesis 4: *"Females show less silence behaviour and less defensive silence behaviour compared to men"*

2.3.4 Collaboration behaviours/Supportive behaviours & Unsupportive Behaviours

The terms *supportive* or *collaboration behaviours* and *unsupportive behaviours* are interpreted differently by researchers. In this research the meaning of *supportive* and *unsupportive behaviours* by O'Donovan et al. (2020) is being followed. According to these authors, the following behaviours can be classified as supportive: "Sharing procedures, knowledge and experience, sharing future plans, active listening, use of inclusive language such as "we", agreeing and responding positively or enthusiastically to input, acknowledging achievements, and congratulating and delegating tasks" (O'Donovan et al., 2020, p. 13). Behaviours that are qualified as *unsupportive* are: "Interrupting, discussions within small sub-groups and reacting cold/ignoring a joke" (O'Donovan et al., 2020, p. 13). Women are believed to have a greater ability to listen than males (Briton & Hall, 1995). Men, on the other hand, are prone to interrupt women and pay less attention to the input of their female colleagues (Cox et al., 2006). *Unsupportive behaviours* also include discussions within small subgroups, for instance joking together (O'Donovan et al., 2020, p 13). This type of behaviour, without involving other team members, enhances solidarity between the two team members, although simultaneously creating a social boundary between them and

other team members (Hay, 2000). Literature suggests women to display more collaborative behaviours, thus the following hypotheses are put forward:

Hypothesis 5a: "*Females show more collaborative/supportive behaviours compared to men*"

Hypothesis 5b: "*Females show less unsupportive behaviours compared to men*"

2.3.5 Familiarity behaviours

Following O'Donovan et al.'s codebook (2020, p. 13) the *familiarity behaviours* include "talking about personal, non-work matters" as well as "laughing about a joke". Research is rather scarce with respect to this behaviour at team level. However, we can assume, based on the previous mentioned behaviours, that they work more or less the same. So did two ethnographic studies in UK businesses find that female managers tend to use small talk mixed with business talk strategically in order to create group cohesion (Case & Oedema-Paul, 2015). Other research notes the importance of humor at the workplace (Romero & Cruthirds, 2006) as it influences and reflects group dynamics (Hay, 2000). It is suggested that it can alleviate stress, increase innovation and improve the chance of completing difficult tasks. Moreover, humor is linked to creative thinking and a better organizational culture (Romero & Cruthirds, 2006). Nonetheless, humor is not the same for men and women (Hay, 2000). While men use humor to impress and highlight similarities, women use humor to encourage solidarity (Hay, 2000). In contrary to popular belief that women are humorless creatures (Hay, 2000) the research of Holmes et al. (2001) proves the opposite to be true. The authors found that women actually encourage humor in business meetings. In fact, they found that women-only meetings generate more humor than male-only meetings (Homes et al., 2001). Given the literature found seems to imply that women show more *familiarity behaviours*, the following hypothesis is suggested:

Hypothesis 6: "*Females show more familiarity behaviour compared to men*"

The way in which the above-mentioned hypotheses are going to be tested will be explained in the following section.

3. METHODOLOGY

3.1 Research Design

This research was carried out by coding video observations of teams which were following the Agile way of working. Both a qualitative as well as a quantitative approach was utilised. The use of a mixed-method approach allowed for a coherent and vigorous answer to the research question (Fetters, Curry, & Creswell, 2013). The qualitative approach includes coding of video observations, whereas the quantitative approach includes a frequency count and comparative analyses.

3.2 Data Collection & Sample

The data for this research was collected by the Organisational Behaviour, Change and Consultancy (OBCC) department during the execution of a great project with a Dutch financial organization implementing Agile. The recorded teams work in sprints, which consist of a planning, refinement and retrospective meeting. The analysed data includes four video recordings of retrospective meetings of Agile teams over the course of one sprint which were recorded in 2019 and 2020. The choice for using the retrospective meetings for the analysis was because the team members have worked with each other for the longest

period of time. According to Hackman (2002) there is a positive relationship between performance and the longevity of a team working together (Hackman, 2002). Therefore, the expectation is that the psychological safety is the highest during the retrospective meeting. The exact team composition during the retrospectives as well as the level of effectiveness of each team can be found in Table 1.

The four meetings had a total duration of 268 minutes, ranging from 45 to 102 minutes with an average of 67 minutes per meeting. The four teams consist in total of 27 participants of which 21 were male and 6 were female. The teams recorded consist of five to ten team members of different nationalities, ages and gender. The age of the men ranges from 27 to 58 years old with a mean age of 38.16. The age of the women ranges from 29 to 48 years old with a mean age of 38.17. In case of the males, six different nationalities were present during the meetings, namely Dutch, Thai, Brazilian, Indian, Russian and "Other". In case of the females, two different nationalities were present during the meetings, namely Dutch and English. The meetings were recorded from different angles, assuring a good overview of the happenings. The four teams were chosen based on the number of males and females present during the retrospective meetings. More specifically, two teams with the highest number of women and two teams with the lowest number of women were chosen to allow for a better comparison. Indeed, the reason for this was to see what impact the presence of women has on the psychological safety of a team in order to ultimately answer the question about how observed psychological safety may vary between genders in effective Agile teams. In Table 1. the teams with a low number of females in comparison to the other teams are called "Low" and the teams with a high number of females in comparison to the other teams are called "High".

Table 1. Team Composition & Effectiveness

Team	Category	Males	Females	Effectiveness mean
01001	Low	7	1	5.47
02001	High	6	2	5.85
07001	High	3	3	4.84
12001	Low	5	0	5.9

3.3 Research Instrument

The coding of the video observations was done at the individual level. All team member's psychological safety behaviours were coded by following the O'Donovan et al. (2020) psychological safety codebook. The following dimensions were used during the coding process: *voice behaviours*, *defensive voice behaviours*, *silence behaviours*, *defensive silence behaviours*, *collaboration behaviours*, *learning or improvement behaviours*, *unsupportive behaviours* and *familiarity behaviours*. All of the meetings were coded by two coders. The second coder is likewise a bachelor student, writing thesis research. In order to reach a baseline of agreement and understanding of the coding process, first a video observation was coded outside of the scope of this research for practice. Based on this practice round, it was decided to continue the coding with coding solely the abovementioned dimensions instead of coding all sub-behaviours separately. This allowed for a much smoother and quicker coding process.

The program Observer XT (version 15) was used to code. The coding happened as follows. All of the team members were assigned a number (1,2,3, etc.) and the different behaviour dimensions were marked with a letter (a,b,c, etc.). When a team member expressed a certain type of behaviour, he or she was coded with the according letter during the time of showing the particular behaviour. In case of unclarity, the transcriptions of the

4

meetings were analysed to assign the correct dimensions to each team member. For all the mentioned types of dimensions, the behaviours were counted for the males and females. The meetings have different lengths; therefore, the data was standardised. This was done by dividing the frequencies of the types of behaviours shown by men and women separately by the total number of behaviours shown during the meeting.

To measure team meeting effectiveness, a 4-item scale was self-developed by taking inspiration from Engleberg and Wynn's (2007) and Baran et al.'s (2012) meeting effectiveness scales. Sample items are: 1) "Overall, our meetings are productive" and 2) "The meetings I attend are worth my time". The Cronbach's alpha of the scale was .904.

3.4 Data Analysis
For the qualitative part of the analysis involving video observations and coding, deductive Thematic Analysis was used (Braun & Clarke, 2006). This analysis allows the researcher to interpret qualitative data, in this case the video observations, through a top-down approach, identifying codes and, potentially themes. For this research, codes were applied deductively on the basis of, and corresponding to, the psychological safety behaviours described in the codebook by O'Donovan et al. (2020).

For the quantitative part of the analysis, comparative tests were implemented. Before executing the test, the normality of the data was checked. The normality test can be found in Table 1 in the Appendix. The data turned out not to be normally distributed. Therefore, the differences in frequency of the shown behaviours by men and women were tested with the non-parametric Mann-Whitney U test. The Mann-Whitney U test examines whether two independently sampled groups differ on a continuous variable (McKnight & Najab, 2010). The significance level (p-value) that has been used during testing was p = 0.05. All analyses have been carried out using IBM SPSS 25.

4. RESULTS

In this section the results found are reported per presented hypothesis. In the following subsections the high female and low female teams are being compared.

4.1 Descriptive Statistics
Altogether, men showed 1649 psychological safety behaviours in total, while women showed 462. Bearing in mind that there were 21 males and 6 females in the sample, this equates to 78.5239 and 81.8334 displayed psychological safety behaviours on average by males and females, respectively.

Table 2. Total Frequency of Behaviours All Teams:

	Male	Female	Total
Voice Behaviour	549	203	**752**
Defensive Voice Behaviour	35	16	**51**
Learning or Improvement Behaviour	83	41	**124**
Silence Behaviour	436	79	**515**
Defensive Silence Behaviour	90	10	**100**
Collaboration Behaviour	281	59	**340**
Unsupportive Behaviour	54	10	**64**
Familiarity Behaviour	121	73	**194**
Total	**1649**	**462**	**2140**

Table 3. Frequency of Behaviour per 1 male and per 1 female All Teams:

	Male	Female	Total
Voice Behaviour	26.1429	33.8333	**59.9762**
Defensive Voice Behaviour	1.6667	2.6667	**4.3334**
Learning or Improvement Behaviour	3.9524	6.8333	**10.7857**
Silence Behaviour	20.7619	13.1667	**33.9286**
Defensive Silence Behaviour	4.2857	1.6667	**5.9524**
Collaboration Behaviour	13.381	9.8333	**23.2143**
Unsupportive Behaviour	2.5714	1.6667	**4.2381**
Familiarity Behaviour	5.7619	12.1667	**17.9286**
Total	**78.5239**	**81.8334**	**160.3573**

	Male	Female	Total
Positive behaviour	49.2381	62.6667	**111.9048**
Negative behaviour	29.2857	19.1668	**48.4525**
Total	**78.5238**	**81.8335**	**160.3573**

4.2 Hypotheses Testing
The first hypothesis meant to explore whether there was a difference between males and females in terms of shown psychological safety behaviours overall. The Mann-Whitney U test was applied to test if there were differences in total psychological safety scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 68) and females (Mdn = 78) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 80.5, z = 1.021, p = 0.316 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of overall *observed* psychological safety behaviours.

In order to investigate whether there was a significant difference between the amount of positive and negative behaviour shown by males and females the Mann-Whitney U test was also run. The positive behaviours being *voice behaviour*, *learning or improvement behaviour*, *collaboration behaviour* and *familiarity behaviour* and the negative behaviours being *defensive voice behaviour*, *silence behaviour*, *defensive silence behaviour* and *unsupportive behaviour*.

Similar to the first hypothesis, sub hypothesis 1a and 1b show no significant difference in the amount of positive and negative behaviours shown by men and women. The Mann-Whitney U test was applied to test if there were differences in positive psychological safety scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 34) and females (Mdn = 62.5) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 84, z = 1.226, p = 0.239 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of *observed* positive psychological safety behaviours.

The Mann-Whitney U test was applied to test if there were differences in negative psychological safety scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 21) and females (Mdn = 23) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 66, z = 0.175, p = 0.887 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in

terms of *observed* negative psychological safety behaviours. However, it is worth to mention that when considering the frequency count, males demonstrated 49.2381 positive psychological safety behaviours on average compared to 62.6667 positive psychological safety behaviours displayed by females on average. With regards to the shown negative psychological safety behaviours, the number on average shown by males is 29.2857 while for women this number is almost half, namely 14.333 negative psychological safety behaviours shown per female. The results from the frequency count suggest thus that females on average present 13.4286 more positive psychological safety behaviours during a retrospective meeting, whilst also showing 14.9524 less negative psychological safety behaviours averagely.

The second hypothesis meant to investigate whether there was a significant difference in *voice behaviour* and *defensive voice behaviour* shown by men and women, hypothesising that females shows less of both behaviours. The Mann-Whitney U test was applied to test if there were differences in *voice behaviour* and *defensive voice behaviours* scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 19) and females (Mdn = 36.5) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 83.5, z = 1.197, p = 0.239 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of *observed voice behaviours* and *defensive voice behaviours*. As can be found in Table 3 males on average show 26.1429 voice behaviours and 1.6667 *defensive voice behaviours* while females show on average 33.8333 *voice behaviours* and 2.6667 *defensive voice behaviours*. Resulting in a difference of 7.6904 and 1, respectively.

The third hypothesis looked into the difference in *learning or improvement behaviours* shown by men and women assuming that females display less of this type of behaviour. The Mann-Whitney U test was applied to test if there were differences in *learning or improvement behaviour* scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 3) and females (Mdn = 4) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 78.5, z = 0.911, p = 0.376 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of *observed learning or improvement behaviours*. Although no significant difference was found, the frequency count shows that, on average, women display more *learning or improvement behaviours* than men. The corresponding average *of learning or improvement behaviours* are 3.9524 for men and 6.8333 for women, which can also be found in Table 3.

Table 3. Significance level of the hypotheses

Hypotheses	Significance level
1	0.512[1]
1a	0.239[1]
1b	0.755[1]
2	0.316[1]
3	0.376[1]
4	0.798[1]
5a	0.670[1]
5b	0.589[1]
6	0.004[1]

The fourth hypothesis meant to explore whether there was a significant difference between men and women with regards to the shown *silence behaviour* and *defensive silence behaviour*, assuming that women show less of both behaviours according to the theory. The Mann-Whitney U test was applied to test if there

were differences in *silence behaviours* and *defensive silence behaviours* scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 18) and females (Mdn = 14) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 58, z = -0.293, p = 0.798 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of *observed* silence *behaviours* and *defensive silence behaviours*. Even though the difference is not significant, on average females show rather less *silence behaviour* compared to men. Men show 20.7619 *silence behaviours* per person, while women show 13.1667 *silence behaviours* on average. With regards to the *defensive silence behaviours*, men also display more of this behaviour, that is 4.2857, while women display 1.6667 *defensive silence behaviours* averagely.

The fifth hypothesis, split up in hypothesis 5a and 5b, was tested to see whether there was a significant difference between the amount of shown *collaboration behaviours* and *unsupportive behaviours* by men and women. The Mann-Whitney U test was applied to test if there were differences in *collaboration behaviours* scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 7) and females (Mdn = 6.5) was not statistically significant, U(N_{males} = 21, $N_{females}$ = 6) = 71, z = 0.468, p = 0.67 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of *observed collaboration behaviours*. The Mann-Whitney U test was applied to test if there were differences in *unsupportive behaviours* scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 2) and females (Mdn = 1.5) was not statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 53, z = -0.593, p = 0.589 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that men and women do not differ in terms of *observed unsupportive behaviours*. When it comes to the frequency count, males showed 13.381 *collaboration behaviours*. This being the only positive psychological safety behaviour in which men score higher than women. When it comes to the *unsupportive behaviour*, men showed 2.5714 *unsupportive behaviours* on average. In case of the women the average number of *collaboration behaviours* was 9.8333, while the average number of *unsupportive behaviours* shown was on average less compared to men, that is 1.6667.

The last, sixth hypothesis, investigated whether there was a significant difference between males and females with regards to the manifested *familiarity behaviours*, assuming that females show more of this type of behaviour. The Mann-Whitney U test was applied to test if there were differences in *familiarity behaviours* scores between males and females. The Mann-Whitney U test indicated that the difference for males (Mdn = 5) and females (Mdn = 12) was statistically significant, U (N_{males} = 21, $N_{females}$ = 6) = 109.5, z = 2.732, p = 0.004 (<0.05). Therefore, since we can reject the null hypothesis, this implies that men and women do differ in terms of *observed familiarity behaviours*. Females indeed show significantly less *unsupportive behaviours* compared to men. When looking at the frequency count, men display this behaviour 5.7619 times on average, whilst women show more than double of the amount of this positive psychological safety behaviour, namely 12.1667.

4.3 Comparison between Teams with a High vs Low Number of Women

The explored teams were categorised in High-Number-of-Females in Team (i.e., team 02001 and team 07001) and Low-Number-of-Females in Team (i.e., team 01001 and team 12001) as was already explained in Table 1. In order to examine whether there was a significant difference in terms of effectiveness between those teams, a Mann-Whitney U test was run. Before executing the test, the normality of the data was checked. The data turned out to not be normally distributed.

The Mann-Whitney U test indicated that the difference for High-Number-of-Females-in-Team (Mdn = 5.3) and Low-Number-of-Females in Team (Mdn = 5.7) was not statistically significant, $U(N_{\text{High-Number-of-Females in Team}} = 2, N_{\text{Low-Number-of-Females in Team}} = 2) = 1.0$, z = -0.775, p = 0.667 (>0.05). Therefore, since we cannot reject the null hypothesis, this implies that a high or low number of women in a team does not influence the team effectiveness.

4.3.1 High-Number-of-Women in Team

The following section more closely considers the results of the teams which have the highest number of females present during the retrospective meeting. Team 02001 and team 07001 had an attendance of 2 and 3 females, respectively. There were 6 men present in team 02001 and 3 men in team 07001. Starting with team 02001, first of all, the teams' effectiveness level amounted to a mean of 5.85. Considering Table 4, the women in this team showed more *voice behaviour*, *learning or improvement* behaviour and more *familiarity behaviour* compared to men as well as less *defensive voice behaviour*, *silence behaviour*, *collaboration behaviour* and *unsupportive behaviour*. Hence, the women in this team showed, on average, more positive psychological safety behaviour in three out of the four positive psychological safety behaviours, as well as less negative psychological safety behaviours in all four negative psychological safety behaviours.

Table 4. Frequency of Behaviour per 1 male and per 1 female Team 02001:

	Male	Female	Total
Voice Behaviour	32.1667	35.5	**67.6667**
Defensive Voice Behaviour	1.6667	1.5	**3.1667**
Learning or Improvement Behaviour	5.6667	6.6667	**12.3334**
Silence Behaviour	23.1667	13	**36.16667**
Defensive Silence Behaviour	2	0.3333	**2.3333**
Collaboration Behaviour	9	6.5	**15.5**
Unsupportive Behaviour	2.5	0.5	**3**
Familiarity Behaviour	9	10.5	**19.5**
Total	**85.1668**	**74.5**	**159.6668**

	Male	Female	Total
Positive behaviour	55.8333	62.5	**118.3333**
Negative behaviour	29.3333	15.5	**44.8333**
Total	**85.1666**	**78**	**163.1666**

Continuing with the results of team 07001. This team being an equally mixed group (3 males and 3 females). In this team females expressed a lot more *voice behaviours* on average as well as more *learning or improvement behaviours*, *collaboration behaviours* and *familiarity behaviours* in the positive psychological safety behaviours category. However, although they did show less *silence behaviours* averagely, the females in this team also showed more *defensive voice behaviours* and a little bit more *defensive silence behaviours* and *unsupportive behaviours* on average in the negative psychological safety behaviours category.

Table 5. Frequency of Behaviour per 1 male and per 1 female Team 07001:

	Male	Female	Total
Voice Behaviour	17.6667	41.6667	**59.3334**
Defensive Voice Behaviour	1	4.3333	**5.3333**
Learning or Improvement Behaviour	3	7	**10**
Silence Behaviour	25	17	**42**
Defensive Silence Behaviour	2.3333	3	**5.3333**
Collaboration Behaviour	6.6667	13.6667	**20.3334**
Unsupportive Behaviour	0.6667	1.6667	**2.3334**
Familiarity Behaviour	4.6667	11	**15.6667**
Total	**61.0001**	**99.3334**	**160.3335**

	Male	Female	Total
Positive behaviour	32.0001	73.3334	**105.3333**
Negative behaviour	29	26	**55**
Total	**61.0002**	**99.3334**	**160.3333**

4.3.2 Low-Number-of-Women in Team

Team 01001 and 12001, with respectively 1 and 0 women present during the meetings, are considered to be the low number of female teams in this research. Regarding the males, team one consists of 7 males and team 12001 consists of 5 males. Despite the small number of women present during the retrospective meeting, the influence of a single woman with regard to psychological safety is greatly noticeable. Analysing the results of team 01001, it is evident that the only female in the team manifested almost as much *voice behaviours* on average as her male team members, as can be seen in Table 6. In the negative psychological safety behaviours category, she expressed, on average, more *unsupportive behaviours*, but less *defensive voice behaviours*, *silence behaviours* and *defensive silence behaviours*. In terms of positive psychological safety behaviours, she displayed, on average, less *learning or improvement behaviours*, but more *collaboration behaviours* as well as *familiarity behaviours*.

Table 6. Frequency of Behaviour per 1 male and per 1 female Team 01001:

	Male	Female	Total
Voice Behaviour	7.5714	7	**14.5714**
Defensive Voice Behaviour	0.4286	0	**0.4286**
Learning or Improvement Behaviour	0.5714	0	**0.5714**
Silence Behaviour	3.4286	2	**5.4286**
Defensive Silence Behaviour	0.2857	0	**0.2857**
Collaboration Behaviour	3.8571	5	**8.8571**
Unsupportive Behaviour	2.1429	4	**6.1429**
Familiarity Behaviour	5.7143	19	**24.7143**
Total	**24**	**37**	**61**

	Male	Female	Total
Positive behaviour	17.7143	31	**48.7143**
Negative behaviour	6.2857	6	**12.2857**
Total	**24**	**37**	**61**

Finishing with the last team, 12001, which was the only team with no females present in the retrospective meeting. The results show the highest average of *voice behaviour* expressed by males out of all the groups while also having the highest frequency on average of shown *silence behaviour* and *defensive silence behaviour*. However, the amount *collaboration behaviour* shown is also the highest among the four groups. Furthermore, it is striking that the *familiarity behaviour* is quite low on average.

Table 7. Frequency of Behaviour per 1 male Team 12001:

	Male	Female	Total
Voice Behaviour	50	0	**50**
Defensive Voice Behaviour	3.8	0	**3.8**
Learning or Improvement Behaviour	7.2	0	**7.2**
Silence Behaviour	39.6	0	**39.6**
Defensive Silence Behaviour	13.8	0	**13.8**
Collaboration Behaviour	36	0	**36**
Unsupportive Behaviour	4.4	0	**4.4**
Familiarity Behaviour	2.6	0	**2.6**
Total	**157.4**	**0**	**157.4**

	Male	Female	Total
Positive behaviour	95.8	0	**95.8**
Negative behaviour	61.6	0	**61.6**
Total	**157.4**	**0**	**157.4**

5. DISCUSSION

The aim of this thesis was to innovatively explore psychological safety behaviours and investigate how they could differ between men and women. By doing this research, this thesis contributed to the psychological safety literature in two ways.

5.1 Theoretical Implications

Firstly, by investigating *observed* psychological safety through video observations, we address the call for more objective measurements that goes beyond self-reported survey (Zhao et al, 2019). The findings of this thesis thus extend current knowledge on *perceived* psychological safety by considering the behaviours associated with it. Secondly, in Agile environments the combination of observed psychological safety and gender has hardly been investigated. Because Agile teams are becoming more popular (Dikert et al., 2016), it is critical to investigate how psychological safety and gender play a role in these teams, as this may help to better understand their impact on team effectiveness and their success.

More specifically, according to the literature, women were supposed to show types of behaviours less (*voice behaviours*, *defensive voice behaviours*, *learning or improvement behaviours*, *silence behaviours*, *defensive silence behaviours* and *unsupportive behaviours*) or more (*collaboration behaviours* and *familiarity behaviours*) frequently compared to men (Le Pine & Van Dyne, 1998; Eibl et al., 2020; Roussin et al., 2018; Heilman, 2012; Phutela 2015; Cox et al., 2006; Briton & Hall, 1995; Case & Oedema-Paul, 2015; Holmes et al., 2001) . Thus, the literature was indicating that there would be a difference. However, our results do not corroborate previous research since they were non-significant for all the hypotheses but the last one (i.e., *familiarity behaviours*). Hence, our findings point to no significant difference between men and women with regard to *observed* psychological safety. This means that even if in a team there are (a lot) less females than males, which is the cases in the financial sector as well as in Agile teams (Fair Bank Guide, 2020; Warnert, 2015), they actually show as many psychological safety behaviours as men.

Moreover, females even showed more behaviours in three out of four positive psychological safety behaviours *(voice behaviour, learning or improvement behaviour* and *familiarity behaviou*r) as well as showed less behaviours in three out of four negative psychological safety behaviours (*silence behaviour, defensive silence behaviour* and *unsupportive behaviour*). This seems to imply that women do have a positive influence on the level of the overall psychological safety climate in a team. This is in line with previous research as it confirms that women have greater listening skills (Briton & Hall, 1995), therefore interrupting less, as well as showing less aggressive behaviours (Phutela, 2015), thus showing less *defensive silence behaviours*. Additionally, this thesis corroborates studies showing that women indeed display more *familiarity behaviours* by using small talk mixed with business talk strategically in order to create group cohesion (Case & Oedema-Paul, 2015). It also confirms Homes et al.'s (2001) research that women encourage humor in business meetings (Homes et al., 2001). Lastly, the results refute the theories suggesting that women would be hesitant to speak up and limit their ideas and suggestions in male-dominated environments (Eibl et al., 2020; Le Pine & Van Dyne, 1998) and confirms Morrison et al.'s (2011) results that there is no relationship between gender and voice.

The above results and their inconsistency with previous literature could be due to several reasons. First of all, the novel approach of this research could have played a role in the non-significant results obtained in this research. This thesis followed the innovative way of measuring psychological safety behaviours through video observations rather than self-reported measures like surveys. The latter are often considered to be subjective and prone to social desirability or approval biases, whereas observations of actual verbal behaviours are considered to be more objective (Zhao et al., 2019). This allowed us to avoid the discrepancy between how the team members felt and how they actually acted (Manfredo & Shelby, 1988). Second, not all

differences in psychological safety behaviors can be attributed to gender. There are other factors at stake that can play a role in how a person behaves during a meeting, for example personality traits (George, 1990) or cultural differences (Cauwelier et al., 2016; Thorgren & Caiman, 2019). The former can indeed influence people's intrinsic nature to, for instance, speak up; the latter can explain the acquired socio-cultural predispositions and preferences of individuals to express their opinions.

Lastly, regarding team effectiveness and whether this differs between high-number-of-females in teams and low-number-of-females in teams, the results of this thesis indicate that gender composition does not have an influence on a team's effectiveness. Albeit scantly researched, this seems to contradict recent research pointing to a positive relationship between agile women leadership and team effectiveness (Akkays & Bagieńska, 2022). This could be due to the fact that the women in the teams considered in this thesis were not in position of leadership or, indeed, to the verbal behaviour approach implemented. Yet, more studies are needed to better understand this relationship further.

5.2 Practical Implications

Not only are the findings of this research valuable for theory, but they are also of use in practice. Given the fact that the Agile method is quite new, but gaining popularity (Dikert et al., 2016) and gender and psychological safety has not been explored much till date in this particular combination, this research adds new insights in this respect. First of all, research has noted women are rather scarce in the financial sector as well as in the Agile environment (Warnert, 2015; Fair Bank Guide, 2020). Considering the positive effect of women's presence in Agile teams, HR managers as well as marketing managers could keep this in mind and decide to reach out to more women in this field. This will not only create more psychological safety within a team, but it will also start a wave of breaking the established patterns of women being a minority in the abovementioned sectors. Additionally, as research has shown the gender composition of a team matters (Woolley et al., 2010; Ivanova-Stenzel & Kübler, 2011; Apesteguia et al, 2012), managers could take the results of this thesis into account when deciding the gender composition of (Agile) teams, especially given the positive and beneficial influence of women in regard to the manifestation of psychological safety behaviours. Thus, in order to let (Agile) teams fully flourish, it is wise to take into consideration the gender composition of a(n) (Agile) team.

6. LIMITATIONS AND FUTURE RESEARCH

There are certain limitations to this thesis that should be considered. Firstly, the limited sample size: only four teams were analysed which could have influenced the outcomes of this study. In order to understand the role that gender plays on psychological safety better, it is advisable to analyse similar topics accounting for a bigger sample size. Secondly, this thesis focused on retrospective meetings: coding the planning and refinement meetings would possibly make some behaviours come forward more frequently and evidently due to the different goals of these meetings, thus allowing for a broader view of the development of the team members' behaviours throughout the whole sprint. Thirdly, overall, there were a few women present in the teams. Therefore, future research could think of involving teams with more than three women as well as teams with female dominant team members, and with all-female team members as results may differ. Fourthly, the codebook used was still in its early stages of development. Although it acted as a great and useful guidance for coding, some elements needed some

improvement. For example, it was not always clear when a team member would exhibit a behaviour that could be classified as *voice behaviour* or as *learning or improvement behaviour*. Among other things, this led to a rather low intercoder reliability rate with an agreement rate ranging from 21% to 41%. Therefore, future research could think of implementing an improved version of the codebook in order to reach a higher agreement rate between the coders.

Fifthly, the duration of the behaviours was not taken into account. Even though the coding process was carried out very precisely by taking into account every detail in seconds, making the number of behaviours expressed by men and women accurate, it could have been interesting to see whether certain behaviours were shown for longer periods by men or by women. Thus, future research could consider implementing the duration of the behaviours and explore whether a difference can be found there when it comes to gender. Lastly, it has not been tracked towards who the expressed behaviour was being expressed/put at. It would be interesting for future research to also look more into the receiving party to look at whether men or women receive more positive and negative psychological safety behaviour, as well as the behaviour expressed to a group as a whole and which gender does this more often.

7. CONCLUSION

In conclusion, although their brains work differently, men and women are not polar opposites of each other as some studies tended to suggest (Eibl et al., 2020; Le Pine & Van Dyne, 1998; Roussin et al., 2018). By expanding the current knowledge on gender in relation to observed psychological safety (i.e., through the innovative approach of the video observations), this research has shown that there is no significant difference in the number of psychologically safety behaviours men and women show. However, what this research did show is that women seem to bring out the best in their teammates by displaying more positive psychological safety behaviours and less negative psychological safety behaviours. Therefore, the results of this thesis suggest that, on the one hand, women, no matter how many in a team, can foster team dynamics by promoting beneficial psychologically safe behaviours; and, on the other hand, that women are outspoken, encourage improvement, as well as enhance the psychological safety atmosphere with their humor within the team.

8. ACKNOWLEDGEMENTS

9. REFERENCES

Abrudan, D., & Conea-Simiuc, I. (2019). How Managers Can Create Happy Working Environment. *Revista de Management Comparat International*, *20*(3), 286-296.

Acharya, B., & Colomo-Palacios, R. (2019, July). A systematic literature review on autonomous agile teams. In *2019 19th International Conference on Computational Science and Its Applications (ICCSA)* (pp. 146-151). IEEE.

Akkaya, B., & Bagieńska, A. (2022). The Role of Agile Women Leadership in Achieving Team Effectiveness through Interpersonal Trust for Business Agility. *Sustainability*, *14*(7), 4070.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293-312.

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, *58*(1), 78-93.

Baran, B. E., Shanock, L. R., Rogelberg, S. G., & Scott, C. W. (2012). Leading group meetings: Supervisors' actions, employee behaviors, and upward perceptions. *Small Group Research*, *43*(3), 330-355.

Bear, J. B., & Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdisciplinary science reviews*, *36*(2), 146-153.

Begel, A., & Nagappan, N. (2007, September). Usage and perceptions of agile software development in an industrial context: An exploratory study. In *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)* (pp. 255-264). IEEE.

Bowles, H. R., & Gelfand, M. (2010). Status and the evaluation of workplace deviance. *Psychological Science*, *21*(1), 49-54.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Briton, N. J., & Hall, J. A. (1995). Beliefs about female and male nonverbal communication. *Sex Roles*, *32*(1), 79-90.

Brown, B., & Anthony, S. D. (2011). How P&G tripled its innovation success rate. *Harvard Business Review*, *89*(6), 64-72.

Burns, L., Marra, M., & Holmes, J. (2001). Women's humour in the workplace: a quantitative analysis. *Australian Journal of Communication*, *28*(1), 83-108.

Case, S. S., & Oetama-Paul, A. J. (2015). Brain biology and gendered discourse. *Applied Psychology*, *64*(2), 338-378.

Cauwelier, P., Ribière, V. M., & Bennet, A. (2016). Team psychological safety and team learning: A cultural perspective. *The Learning Organization*.

Chaudhuri, A., & Gangadharan, L. (2002). *Gender differences in trust and reciprocity* (No. 2002-03). Wellesley College Working Paper.

Cox, J. (2006). *Understanding doctors' performance*. Radcliffe Publishing.

Díaz-García, C., González-Moreno, A., & Jose Sáez-Martínez, F. (2013). Gender diversity within R&D teams: Its impact on radicalness of innovation. *Innovation*, *15*(2), 149-160.

Dikert, K., Paasivaara, M., & Lassenius, C. (2016). Challenges and success factors for large-scale agile transformations: A systematic literature review. *Journal of Systems and Software*, *119*, 87-108.

Dyne, L. V., Ang, S., & Botero, I. C. (2003). Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of management studies*, *40*(6), 1359-1392.

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological bulletin*, *108*(2), 233.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), 350-383.

Edmondson, A. C. (2002). *Managing the risk of learning: Psychological safety in work teams* (pp. 255-275). Cambridge, MA: Division of Research, Harvard Business School.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual review of organizational psychology and organizational behavior*, *1*(1), 23-43.

Eliot, L. (2011). The trouble with sex differences. *Neuron*, *72*(6), 895-898.

Engleberg, I. N., & Wynn, D. R. (2007). *Working in groups* (4th ed.). Boston, MA: Houghton Mifflin

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs - Principles and practices. Health Services Research, 48(6), 2134- 2156.

Fowler, M., & Highsmith, J. (2001). The agile manifesto. *Software development*, *9*(8), 28-35.

George, J. M. (1990). Personality, affect, and behavior in groups. *Journal of applied psychology*, *75*(2), 107.

Grzeszczyk, T. A. (2020). Development of human capital in institutions from the Polish financial sector: Towards new technologies and agile business models. *International Entrepreneurship Review*, *6*(4), 51-63.

Hackman, J. R., & Hackman, R. J. (2002). *Leading teams: Setting the stage for great performances*. Harvard Business Press.

Halliday, C. S., Paustian-Underdahl, S. C., Stride, C., & Zhang, H. (2022). Retaining Women in Male-Dominated Occupations across Cultures: The Role of Supervisor Support and Psychological Safety. *Human Performance*, 1-22.

Hay, J. (2000). Functions of humor in the conversations of men and women. *Journal of pragmatics*, *32*(6), 709-742.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior*, *32*, 113-135.

Hoda, R., Noble, J., & Marshall, S. (2012). Self-organizing roles on agile software development teams. *IEEE Transactions on Software Engineering*, *39*(3), 422-444.

Hoeseb, C. H., & Tanner, M. (2020, December). Large-Scale Agile Implementation in Large Financial Institutions: A Systematic Literature Review. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1780-1786). IEEE.

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, *59*(7), 1514-1528.

Ivanova-Stenzel, R., & Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, *32*(5), 797-808.

Lam, C. F., & Mayer, D. M. (2014). When do employees speak up for their customers? A model of voice in a customer service context. *Personnel Psychology*, *67*(3), 637-666.

Lee, H. W., Choi, J. N., & Kim, S. (2018). Does gender diversity help teams constructively manage status conflict? An evolutionary perspective of status conflict, team psychological safety, and team creativity. *Organizational Behavior and Human Decision Processes*, *144*, 187-199.

LePine, J. A., & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of applied psychology*, *83*(6), 853.

Manfredo, M. J., & Shelby, B. (1988). The Effect of Using Self-Report Measures in Tests of Attitude—Behavior Relationships. *The Journal of Social Psychology*, *128*(6), 731-743.

McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. *The Corsini encyclopedia of psychology*, 1-1.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science*, *317*(5834), 82-82.

Melnik, G., & Maurer, F. (2006, June). Comparative analysis of job satisfaction in agile and non-agile software development teams. In *International conference on extreme programming and agile processes in software engineering* (pp. 32-42). Springer, Berlin, Heidelberg.

Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management annals*, *5*(1), 373-412.

Morrison, E. W., & Milliken, F. J. (2000). Organizational silence: A barrier to change and development in a pluralistic world. *Academy of Management review*, *25*(4), 706-725.

Morrison, E. W., Wheeler-Smith, S. L., & Kamdar, D. (2011). Speaking up in groups: a cross-level study of group voice climate and voice. *Journal of Applied Psychology*, *96*(1), 183.

Nicholls, A. R., Polman, R., Levy, A. R., Taylor, J., & Cobley, S. (2007). Stressors, coping, and coping effectiveness: Gender, type of sport, and skill differences. *Journal of sports sciences*, *25*(13), 1521-1530.

O'Donovan, R., Van Dun, D., & McAuliffe, E. (2020). Measuring psychological safety in healthcare teams: developing an observational measure to complement survey methods. *BMC medical research methodology*, *20*(1), 1-17.

Ortmann, A., & Tichy, L. K. (1999). Gender differences in the laboratory: evidence from prisoner's dilemma games. *Journal of Economic Behavior & Organization*, *39*(3), 327-339.

Phutela, D. (2015). The importance of non-verbal communication. *IUP Journal of Soft Skills*, *9*(4), 43.

Pinder, C. C., & Harlos, K. P. (2001). Employee silence: Quiescence and acquiescence as responses to perceived injustice. In *Research in personnel and human resources management*. Emerald Group Publishing Limited.

Przybilla, L., Wiesche, M., & Krcmar, H. (2019, June). Emergent Leadership in Agile Teams--an Initial Exploration. In *Proceedings of the 2019 on Computers and People Research Conference* (pp. 176-179).

Rogelberg, S. G., & Rumery, S. M. (1996). Gender diversity, team decision quality, time on task, and interpersonal cohesion. *Small group research*, *27*(1), 79-90.

Romero, E. J., & Cruthirds, K. W. (2006). The use of humor in the workplace. *Academy of management perspectives*, *20*(2), 58-69.

Roussin, C. J., Larraz, E., Jamieson, K., & Maestre, J. M. (2018). Psychological safety, self-efficacy, and speaking up in interprofessional health care simulation. *Clinical Simulation in Nursing*, *17*, 38-46.

Schulte-Rüther, M., Markowitsch, H. J., Shah, N. J., Fink, G. R., & Piefke, M. (2008). Gender differences in brain networks supporting empathy. *Neuroimage*, *42*(1), 393-403.

Scott, E., Milani, F., Kilu, E., & Pfahl, D. (2021). Enhancing agile software development in the banking sector—A comprehensive case study at LHV. *Journal of software: evolution and process*, *33*(7), e2363.

Seong, J. Y., & Hong, D. S. (2013). Gender diversity: How can we facilitate its positive effects on teams?. *Social Behavior and Personality: an international journal*, *41*(3), 497-507.

Takeda, S., & Homberg, F. (2014). The effects of gender on group work process and achievement: an analysis through self- and peer-assessment. *British Educational Research Journal*, *40*(2), 373-396.

The Centre for Research on Multinational Corporations (SOMO). (2020). *From glass ceilings to factory floors*. SOMO. https://www.somo.nl/from-glass-ceilings-to-factory-floors/

Thorgren, S., & Caiman, E. (2019). The role of psychological safety in implementing agile methods across cultures. *Research-Technology Management*, *62*(2), 31-39.

Warnert, N. A. (2015). *Increasing Women's Involvement in the Agile Community* (Doctoral dissertation, St. Catherine University).

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, *330*(6004), 686-688.

Zaidi, Z. F. (2010). Gender differences in human brain: a review. *The open anatomy journal*, *2*(1).

Zhao, E. Y., Thatcher, S. M., & Jehn, K. A. (2019). Instigating, engaging in, and managing group conflict: A review of the literature addressing the critical role of the leader in group conflict. *Academy of Management Annals*, *13*(1), 112-147.

10. APPENDIX

Appendix 1. Normality Test

	Tests of Normality					
	Kolmogorov-Smirnov[a]			*Shapiro-Wilk*		
	Statistic	*df*	*Sig.*	*Statistic*	*df*	*Sig.*
Voice Behaviour	0.180	27	0.024	0.863	27	0.002
Defensive Voice Behaviour	0.305	27	0.000	0.539	27	0.000
Learning or Improvement Behaviour	0.214	27	0.003	0.831	27	0.000
Silence Behaviour	0.193	27	0.011	0.825	27	0.000
Defensive Silence Behaviour	0.368	27	0.000	0.507	27	0.000
Collaboration Behaviour	0.250	27	0.000	0.736	27	0.000
Unsupportive Behaviour	0.178	27	0.028	0.839	27	0.001
Familiarity Behaviour	0.139	27	0.193	0.921	27	0.043

a. Lilliefors Significance Correction