Observed Psychological Safety and Individual Job Performance: Behavioral Differences in Monocultural and Multicultural Agile Teams

Author: Mette Sophie in 't Anker
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

Given today's globalized world, the importance of multicultural teams has significantly increased over the past decades. However, the composition of these teams, especially when working Agile, can potentially impact psychological safety during meetings, which in turn might affect individual job performance. Therefore, this thesis aims to explore how *observed* psychological safety can differ between mono- and multicultural teams and its impact on individual job performance. To be able to obtain more objective and reliable results, psychological safety behaviors were *observed* during Sprint Retrospective meetings of two monocultural and two multicultural agile teams, and related to individual job performance. Consequently, an exploratory sequential study was conducted through a mixed-method research design, with both qualitative and quantitative analyses. The findings of this thesis underline that monocultural agile teams seem to have slightly higher levels of observed psychological safety in their meetings than multicultural agile teams. This is likely due to a lower presence of psychologically unsafe behaviors (such as defensive voice behaviors, defensive silence behaviors, and unsupportive behaviors) in monocultural agile teams. Furthermore, a positive relationship has been established between observed psychological safety and individual job performance in the episode analysis, and in one comparison of the comparative frequency analysis. To increase individual job performance, the practical implications of this thesis suggest that organizations should increase the cultural knowledge and embracement of other cultures, and increase awareness about the effects of psychological safety on individual job performance to the employees.

Graduation Committee members:
Dr. L. Carminati
Dr. P. Weritz

Keywords

Agile project management, Behavioral differences, Job performance, Monocultural teams, Multicultural teams, Psychological safety

1. INTRODUCTION

Over the past decades, many businesses have faced an environment that is increasingly competitive and is characterized by continuous and unpredictable change (Christopher, 2000). Due to this fast pace of change, the agile way of working originated. Agile project management approaches contrast with traditional approaches (such as waterfall) by highlighting 'continuous design, flexible scope, freezing design features as late as possible, embracing uncertainty and customer interaction, and a modified project team organization' (Serrador & Pinto, 2015, p. 1041). In the agile way of working, agile teams, also called squads, operate as self-managed, cross-disciplinary teams that tend to rely on shared leadership (Šmite et al., 2023). However, since agile is a relatively new concept in many different industries, several challenges relating to how employees cope with agile may arise (Zielske & Held, 2022).

One of these challenges is related to promoting psychological safety. The extent to which team members share their knowledge and speak up with suggestions and opinions (Mesmer-Magnus & DeChurch, 2009), also referred to as employee voice (Morrison, 2014), is maximized in a psychologically safe environment. It is often seen that employees keep their suggestions and opinions to themselves, where psychological barriers the employees face are one of the key factors observed (Morrison, 2014). That is why in order for employees to use their voice, it is increasingly important to create an environment in agile teams that contributes positively to psychological safety. Indeed, psychological safety was identified as the number one characteristic of successful high-performing teams (Bergmann & Schaeppi, 2016), and contributing positively to individual job performance (Miao et al., 2019), individual effectiveness, and team effectiveness (Newman et al., 2017)

Furthermore, given today's globalized world, psychological safety plays a crucial role in multicultural work teams (Glikson et al., 2016). Multicultural work teams can be defined as groups of people 'from different cultures, with a joint deliverable for the organization or another stakeholder' (Stahl et al., 2010, p. 439), and their importance has significantly increased over the past decades. (Groves & Feyerherm, 2011). According to Glikson et al. (2016), whilst multicultural teams tend to have a global identity that can be positively linked to a psychologically safe work environment, monocultural teams, where members only have one nationality (Leifels & Bowen, 2021), do not have this global identity. Global identity 'allows to perceive culturally different others as an in-group, and thus lowers the levels of identity threat and increases the sense of safety' (Glikson et al., 2016, p. 5). Hence, monocultural teams are less likely to show psychologically safe behavior. On the contrary, research done by Thorgren and Caiman (2019, p. 31) suggested that there are cultural differences in agile teams in three areas that may affect workers' perception of psychological safety in the workplace: '(1) cultural differences related to attitudes toward inclusiveness, (2) cultural differences related to perceptions of and trust in collective responsibility, and (3) cultural differences related to openness in communication'. These cultural differences suggest that working across cultural boundaries brings additional challenges, and thus imply that members of multicultural teams may perceive their psychological environment as less safe. Since members of multicultural work teams already face communication challenges related to different perspectives, expectations and language-related barriers (Glikson et al., 2016), it is important that psychological safety is maximized to be able to effectively use employee voice. Also, given the importance of psychological safety related to performance (Bergmann & Schaeppi, 2016) and effectiveness (Newman et al., 2017), psychological safety is crucial for monocultural teams, too. Because the literature is inconsistent in its results regarding psychological safety in mono- and multicultural teams, and given its importance, this relationship needs to be explored more.

Moreover, psychological safety has mostly been studied through survey data (Newman et al., 2017), which may be influenced by self-report biases, thus making the data less reliable (Bauhoff, 2011). Hence, to be able to achieve more objective results, developing an observational measure to complement survey methods is valuable to be explored (O'Donovan et al., 2020). Therefore, to be able to research observed psychological safety through its verbal and nonverbal behavior, this exploratory research *observes* psychological safety during meetings of agile teams. In doing so, this research aims to address the research gap by observing psychological safety through verbal and nonverbal behaviors during meetings of mono- and multicultural agile teams, to be able to relate it to job performance. From this aim, the following research question can be redirected:

How is observed psychological safety displayed in monocultural versus multicultural agile teams, and how does this influence individual job performance?

Thus, this research compares observed psychological behaviors in meetings from monocultural and multicultural agile teams, which are then linked to individual job performance.

This paper contributes to the research done in the field of psychological safety in agile work teams in multiple ways. First, this research is conducted by innovatively observing and comparing behaviors related to psychological safety during retrospective meetings of an agile sprint. Since most studies on psychological safety are done by implementing survey-based methods, this work extends current knowledge on more objective methodological ways to capture psychological safety. Second, this thesis compares observed psychological safety in the unique setting of multi- and monocultural agile teams, and ultimately relates observed psychological safety with job performance. Given the increased importance of agile, globalization and job performance over the past decades (Serrador & Pinto, 2015; Groves & Feyerherm, 2011; Rattini, 2023), understanding whether there is a difference in manifestation of psychological safety between mono- and multicultural agile teams becomes pivotal. Lastly, this paper can also offer practical implications and serve as a recommendation for organizations by providing suggestions on how to address psychological safety when the objective of the organization is to maximize job performance.

In the remainder of this thesis, the theoretical framework behind the research question will be explained, continued by the methodology used to conduct the research. Furthermore, the results will be presented, which will then be discussed. A conclusion can be found at the end of the paper, supported by recommendations for further research.

2. THEORETICAL FRAMEWORK

2.1 Agile teams

According to Cooper and Sommer (2016, p. 514), 'agile is a project management method that brings agility, adaptability, and speed to development projects: It includes micro-planning tools for creating software code and get to a working end-product quickly'. Agile teams are working towards four core values: 1) Individuals and interactions over processes and tools, 2) Working software over comprehensive documentation, 3) Customer collaboration over contract negotiation, and 4) Responding to change over following a plan (Dybå & Dingsøyr, 2008). These values are part of the Agile Manifesto, based on how traditional project management approaches view their way of working vs.

how agile project management would like to work. Agile teams are self-managed (Šmite et al., 2023), where most agile teams consist of a product owner (Lindsjørn et al., 2016) and other specialized team members, each in different disciplines (Šmite et al., 2023).

In general, agile projects consist of a number of very short sprints, also known as iterations, where each sprint or iteration produces an executable finished product (Cooper & Sommer, 2016) in one to four weeks (Grapenthin et al., 2015). One of the characteristics of a sprint is that it generally has three meetings, where the first meeting is called the Sprint Planning meeting. This meeting is held at the beginning of the sprint, where it is decided what to develop during the sprint (Paasivaara et al., 2009). A second meeting, the Review meeting is scheduled during the sprint, where the established fixed set of tasks in the first meeting is refined and possible results are presented to the stakeholders (Grapenthin et al., 2015). The last meeting, the Sprint Retrospective, is for reflecting on the work done by each team member, with its goal of optimizing future sprints (Grapenthin et al., 2015).

Over recent years, many large organizations have made the decision to fully go from traditional approaches to agile (Paasivaara, 2017). Many frameworks have been developed, under which the Scaled Agile Framework (SAFe), Large-scale Scrum (LeSS), or Disciplined Agile Delivery (DAD) (Paasivaara, 2017), where it is proposed how large organizations can scale organizational agility. Achieving organizational agility is still perceived as a highly complex process (Calnan & Rozen, 2019; Sommer, 2019). However, organizations that are further in the process of transforming to agile, achieve 'around 30% gains in efficiency, customer satisfaction, employee engagement, and operational performance' and are five to ten times faster in their organizational processes than agile competitors that are less far in the transformation (Aghina et al., 2020, p. 2). One of the factors that can help the agile way of working is psychological safety, since this enables employees to independently seize opportunities, adapt, and think (Cai et al., 2018).

2.2 Psychological safety

2.2.1 Perceived psychological safety

Psychological safety can be defined as 'a shared belief that the team is safe for interpersonal risk taking' (Edmondson, 1999, p. 351), where a psychologically safe work environment can be characterized by employees that feel safe to voice ideas, willingly seek and provide honest feedback, collaborate, take risks and experiment (Edmondson, 1999). According to Edmonson (2004), the level of psychological safety within a team depends on the team leader behavior, informal group dynamics, trust and respect, use of practice settings and the existence of a supportive organizational context.

A psychologically safe work environment can be assessed through the level of employee silence and employee voice (Morrison, 2014). Employee voice can be referred to as the extent to which team members share their knowledge and speak up with suggestions and opinions (Morrison, 2014). On the other hand, employee silence is intentionally withholding ideas, information, and opinions with relevance to improvements in work and work organizations (Morrison & Milliken, 2000). In a psychologically safe work environment, employee voice can be maximized, and in a psychologically unsafe work environment, employee silence can be the dominant variable of the two (Van Dyne, Ang & Botero, 2003). Van Dyne et al. (2003) did research in the field of voice behavior, where they were able to divide employee voice and silence into three types: 1) Acquiescent voice and silence, which is expressing or withholding relevant ideas, information, or opinions that are based on *resignation*; 2) Defensive voice and silence, which is expressing or withholding relevant ideas, information, or opinions as a form of self-protection, which is based on *fear*; and 3) Prosocial voice and silence, which is expressing or withholding work-related ideas, information, or opinions with the aim of *benefiting other people or the organization*. These behaviors can be seen in meetings and are also relevant for the performance of agile teams, since employee silence has implications for individual job performance.

2.2.2 Observed psychological safety

In general, psychological safety can be observed through verbal and non-verbal behavior, where the verbal behavior especially links to employee voice (the extent to which team members share their knowledge and speak up with suggestions and opinions; Mesmer-Magnus & DeChurch, 2009), and non-verbal behavior to employee silence (intentionally withholding ideas, information, and opinions with relevance to improvements in work and work organizations; Morrison & Milliken, 2000). These verbal and non-verbal behaviors can, in turn, be characterized by 'supportive, unsupportive, learning or improvement-oriented and familiarity type behaviours' (O'Donovan et al., 2020, p. 1).

Verbal and non-verbal behaviors can be observed during any collaboration between team members (Morrison & Milliken, 2000). However, most research on psychological safety has been done mostly relying on surveys (Newman et al., 2017), which allows to measure perceived psychological safety. As a result, this survey research method can promote self-report biases, obtained from the fact that respondents cannot recall perfectly or provide misleading answers on purpose (Bauhoff, 2011). The verbal and non-verbal behaviors are therefore not precisely taken into consideration in these studies, since employees mostly cannot recall every single interaction during a meeting, while they are of great importance to observe psychological safety (O'Donovan et al., 2020). Consequently, in this thesis observed psychological safety during agile meetings based on the related voice behaviors is considered. This way the self-report biases can be accounted for, and a more objective and reliable evaluation of psychological safety can be attempted.

One of the factors that might influence observed psychological safety is cultural diversity, since different cultures may have impact on the behavior of individuals (Battistella et al., 2023). Because globalization has become increasingly important over the past decades (Groves & Feyerherm, 2011), organizations also need to deal with cultural diversity to an increasing extent.

2.3 Cultural diversity (monocultural and multicultural teams)

Over recent years, multicultural teams have become more common (Ochieng & Price, 2010), since organizations have become increasingly global (Groves & Feyerherm, 2011). Multicultural work teams can be defined as 'a group of people from different cultures, with a joint deliverable for the organization or another stakeholder' (Stahl et al., 2010, p. 439). Project management can be impacted by multicultural teams, since cultures may make people behave differently (Battistella et al., 2023). This can result in additional challenges, as well as opportunities (Stahl et al. 2010), as opposed to monocultural work teams, where members only have one nationality (Leifels & Bowen, 2021). The extent of cultural diversity in a team also has different outcomes, since a more dispersed multicultural team tends to pay more attention to cultural differences, and is less prone to behavior such as cultural stereotyping, forming subgroups based on nationality, or social exclusion of team

members who do not belong to the dominant values of the team (Stahl & Maznevski, 2021).

2.4 Cultural Diversity and Psychological Safety

Recent studies have been trying to explore the relationship between cultural differences and psychological safety. For instance, Leersnyder et al. (2022) studied the concept of psychological safety in international and domestic classrooms. They found that due to cultural misunderstandings and other cultural barriers which are associated with psychological costs, multicultural classrooms experience a lower sense of psychological safety than the monocultural classrooms, especially when students had the feeling that cultural differences were ignored and devalued. However, when cultural differences were valued and embraced, students in the multicultural and monocultural classrooms did not differ in terms of perceived cultural misunderstandings, which resulted in the same level of perceived psychological safety in both monocultural and multicultural contexts.

Similarly, Glikson et al. (2016) found that multicultural team members can feel a certain level of global identity, and when this was at a high level, they were more successful in developing psychological safety. However, high performance was only reached by teams with high collective intelligence, since only they were able to utilize their global identity (Glikson et al., 2016). Collective intelligence is 'a team's ability to collaborate and coordinate effectively, more so than individual team member cognitive ability' (Glikson et al., 2016, p. 3; Kim et al., 2017), which is positively linked to individual and team performance (Glikson et al., 2016). Thus, this research highlights the importance of collective intelligence as a factor of perceived psychological safety, since only then the benefits of the teams' global identity could be utilized, and it highlights the importance of cognitive ability as a positive contribution to individual job performance.

Furthermore, the way team members perceive their other team members also influences the level of psychological safety (Farley et al., 2022). Because team members from an international context are often perceived as less competent than the local ones, international team members develop psychological safety more slowly over time and have lower initial psychological safety to begin with. However, when international team members already had experience with working in multicultural teams, they developed psychological safety more quickly (Farley et al., 2022). Additionally, Farley et al. (2022) found that Europeans have the highest initial psychological safety, followed by UK citizens, Africans and lastly Asians. This is partly because in general African and Asian team members speak up less, which can be linked to their culture (Farley et al., 2022). The fastest growth in psychological safety for international team members from day 1 to 5 in Farley et al. (2022)'s research was for team members from the UK, followed by Africans, Europeans, and Asians respectively.

Lastly, Dibble et al. (2019) added cultural intelligence, also referred to as CQ, as a factor that indirectly influences psychological safety in multicultural teams. Cultural intelligence can be defined as 'a person's capability to adapt effectively to new cultural contexts' (Earley and Ang, 2003, p. 59). The dimensions of CQ consist of (1) an individual's knowledge of another culture, (2) how motivated an individual is to engage with other cultures, (3) an individual's behavior in communicating with others from different cultures, and (4) the individual's cognitive awareness of how others interact with different cultures (Dibble et al., 2019; Earley and Ang, 2003). Whenever the level of CQ was considered high, team members of global teams reflected a sense of improvement in individuals' capability to effectively function in culturally diverse settings, which positively contributes to job performance and psychological safety (Dibble et al., 2019).

When combining the findings of the studies above, it can be noted that cultural diversity seems to have important impact on psychological safety based on the following factors: cultural misunderstandings and barriers, embracement or rejection of cultural differences, global identity, collective and cultural intelligence, and what culture international team members come from. Because cultural diversity seems to have an important impact on psychological safety, the effect of these different levels of a psychologically safe work environment can also be seen in the level of individual job performances per team (Miao et al., 2019; Glikson et al., 2016).

2.5 Psychological Safety and Individual Job Performance

A psychologically safe work environment is particularly important when connecting it to individual job performance of the employees, since there is a positive relationship between the two variables when researched in a monocultural setting (Miao et al., 2019) as well as in a multicultural setting (Glikson et al., 2016). Due to the fact that high levels of psychological safety positively influence team learning, which is a potential resource for an organization in improving its levels of competitiveness in this dynamic and complex environment (Breso et al., 2008), it also positively influences individual job performance (Cauwelier et al., 2016). However, when the level of psychological safety reduces, team learning and individual job performance reduce too. Individual job performance is, in turn, a very important factor for the survival of an organization (Shanafelt & Noseworthy, 2017), and thus important to maximize.

3. METHODOLOGY

3.1 Research design

This exploratory sequential research aims to fill a research gap by trying to find a relationship between cultural diversity and *observed* psychological safety in agile teams, and whether this may influence job performance. To be able to do this, the thesis is based on a mixed-method research design, which consists of a qualitative and quantitative approach (Creswell et al., 2003). Because this is exploratory sequential research, the qualitative components are dominant, and the quantitative components are complementary in this study (Gonzalez-Diaz & Bustamante-Cabrera, 2021). Qualitative research adds great value in obtaining detailed contextualized information, resulting in findings of complex social phenomena. However, to understand these complexities, different kinds of research methods are needed (Creswell et al., 2003). Therefore, this research adds a quantitative approach to be able to create understandable, more in-depth, and detailed findings (Creswell et al., 2003). Due to this triangulation, the results are also more valid (Saunders et al., 2009).

In this thesis, the first qualitative analysis was based on own interpretations (checked with another research to increase reliability in our final dataset) of the observed psychological safety behavior of participants in recorded Sprint Retrospective meetings of an agile sprint. Second, for the quantitative analysis, a comparative frequency analysis was performed on the observed behaviors per team meeting, which were later compared to how frequently other teams' psychological safety behaviors were observed in the meetings. Third, a comparison analysis was performed to find potential differences between the observed psychological safety behaviors and the perceived psychological

safety by the participants based on a questionnaire done throughout the teams. Lastly, another qualitative analysis was performed in the form of an episode analysis, to be able to compare certain episodes of meetings between team members. This analysis added more depth to the frequency analysis findings.

3.2 Sampling and data collection

In collaboration with a Dutch financial service organization, the Organizational Behavior, Change Management and Consultancy (OBCC) Group collected data over a time of 4 years for a project focused on team effectiveness. For this project, multiple data sources were gathered with the help of video recordings of meetings, surveys, arousal equipment and existing databases. Concerning video observations, the three meetings from an agile sprint have been recorded to be able to be observed. In total, the earlier defined Sprint Planning, Sprint Review and the Sprint Retrospective meetings of nine agile teams have been recorded. For this thesis, the focus is on the third video-recorded meetings of the sprint, namely the Sprint Retrospective meeting, since this meeting reflects upon the work done by each team member (Grapenthin, Poggel, Book & Gruhn, 2015). This meeting is most relevant to this research, since this meeting reflects the job performance per team member, from which can be predicted that the emotions will be stronger. The meetings are taken from two monocultural teams and two multicultural teams. Between the two meetings of the mono- and multicultural team a distinction is made between a high performing team opposed to a low performing team. This results in the teams presented in Table 1, a matrix of the four teams researched:

Table 1: a matrix of the four teams researched

	High performance	Low performance
Monocultural	Team 1	Team 2
Multicultural	Team 3	Team 4

The data needed to distinguish the teams were collected through a cross-sectional survey done throughout all teams. The averages of job performances per individual team member were taken per team, to be able to distinguish what teams had high performing team members and what teams had low performing team members.

3.3 Sample description

As indicated in table 1, the sample entails four teams. To avoid long sentences, the teams are called by their number, and not by their specifications. Team 1 consisted of two women (22.2%) and 7 men (77.8%), where the participants had an average age of 35.76 ranging from 22 to 55. Since this is a monocultural team, all members of the team were Dutch. Team 2 only consisted of men (100%), with an average age of 45.2, ranging from 30 to 55. The team consisted of five participants, all Dutch. Team 3 was a multicultural team, where all participants were male (100%). The nationalities represented were Estonian (14.3%), Hungarian (14.3%), Belgian (14.3%) and Dutch (57.1%). The average age was 32.7, ranging from 27 to 48. Lastly, team 4 was a multicultural team consisting of three female (50%) and three male participants (50%). The cultures represented were English (16.67%), Thai (16.67%), Brazilian (16.67%), Russian (16.67%) and Dutch (33.33%). The average age was 41.67 ranging from 37 to 48.

3.4 Measures

3.4.1 Observed Psychological Safety

At the beginning of this research, the four videorecorded meetings were observed and coded using a Psychological Safety Codebook provided by the OBCC Group of the University of Twente. The codebook provides and explains different behaviors that are related to psychologically safe behavior and psychologically unsafe behavior, which can be grouped by the following:

Psychologically safe behaviors: Voice Behaviors, Collaboration Behaviors, Learning or Improvement Oriented Behaviors, and Familiarity Behaviors

Psychologically unsafe behaviors: Defensive Voice Behaviors, Silence Behaviors, Defensive Silence Behaviors, and Unsupportive Behaviors

The psychological safety behaviors were coded by two independent coders, who then compared their interpretations to create one reliable Golden File (i.e. a file of the interpretations of the two coders combined, after agreeing collectively what behavior fits best at what point in time). Before the Golden File was made, the Kappa statistic was calculated for both independent observations to measure interrater reliability, which is the extent to which the coders link the same score to the same behavior (McHugh, 2012). The Kappa value needed to be above .60 to obtain statistical significance (McHugh, 2012), otherwise there were too many possible errors in the data. Considering this requirement, both coders compared their observations and created the Golden File based on how they collectively agreed who's observed behavior fit best at what time. In the end, only Team 3 was coded by two coders, reaching a kappa of .31. The other three teams have only been coded by the researcher.

3.4.2 Individual Job Performance

The individual job performance data have been taken from a cross-sectional survey done throughout all teams, which has been filled in by the leader of the team after the Retrospective Sprint meeting had been finished. This survey consisted of four items, based on a scale made by Gibson et al. (2009): 1) This employee is consistently high performing, 2) This employee is effective, 3) This employee makes few mistakes, and 4) This employee does high quality work. These items were rated on a Likert scale from 1 (very inaccurate) to 7 (very accurate), with reliability $\alpha = 0.86$. The desired α-value is between 0.70 and 0.95 (Tavakol & Dennick, 2011) which indicates that this survey has an acceptable reliability.

3.4.3 Monocultural and multicultural teams

The teams are identified by the definitions given in the Introduction, which means that:

Monocultural teams: Team members only have one nationality (Leifels & Bowen, 2021). Since the sample consists of teams from a Dutch financial service organization, the monocultural teams in question speak Dutch in the meetings.

Multicultural teams: Team members are 'from different cultures, with a joint deliverable for the organization or another stakeholder' (Stahl et al, 2010, p. 439). These teams speak English during the meetings.

The teams were separated into these two categories by observing the spoken language during the meetings. After that, the nationalities of the multicultural teams were also checked through the data previously collected via the cross-sectional survey, to be able to describe the sample of this research.

3.4.4 Perceived psychological safety

Perceived psychological safety was measured through a questionnaire administrated after every meeting. The individual psychological safety was tested by asking the participants in each team to rate three items: 1) During this past meeting, it felt safe for me to make suggestions, 2) During this past meeting, it felt

safe for me to give my opinions, and 3) During this past meeting, it felt safe for me to speak up. The scale used to rate these items was a Likert scale from 1 (strongly disagree) to 7 (strongly agree). In this thesis, the perceived psychological safety from the Retrospective Sprint meeting was used, since all other variables were also measured from this meeting. This survey had a reliability alpha of .90. Since the desired α-value is between 0.70 and 0.95 (Tavakol & Dennick, 2011), an alpha of .90 indicates that this survey has an acceptable reliability.

3.5 Data analysis

3.5.1 Qualitative analysis of videos

To analyze the observed psychological behavior in each selected Retrospective meeting, a deductive version of thematic analysis (Braun & Clarke, 2006) was done to interpret the pattern between the behavior seen and the behavior described in the codebook that the OBCC group provided. The coders started to observe the psychological safety behaviors as soon as the meeting started which was indicated by the door being closed in the meeting room. These behaviors were then observed per individual team member, per second, until the door of the meeting room opened again, which indicated that the meeting ended. A specific behavior was given a start and an end, which indicated the duration of the psychological safety behavior. The behaviors were listed in the software program Observer, to be able to make a file of observed behaviors per individual coder. The individual coders of the meetings then created a Golden File after the interrater reliability was considered high enough (see section 3.4.1).

3.5.2 Quantitative analysis

After all Golden Files of the sample were created, the files were uploaded as a dataset to the software program R-studio, to be able to use descriptive statistics and inferential statistics for comparative analysis.

To be able to compare the datasets, the variables measured in the meetings needed to be standardized. Standardization is crucial, because every meeting had a different number of team members, and a different duration. This resulted in different frequencies and durations of specific behaviors than in other meetings. Therefore, the frequency of behaviors was standardized by dividing the frequency of specific behaviors by the total frequency of behaviors during the meeting, which makes a percentage of specific behaviors measured. Additionally, the duration of each behavior was standardized. The duration of behaviors is the sum of all durations of the specific behaviors observed. This number was standardized by dividing it by the total sum of durations of all behaviors seen in the meeting. After this, again, a percentage was made from this number.

After standardizing, the observed psychological safety was compared in high performing and low performing teams, and in monocultural and multicultural teams. This comparative analysis method was performed on four teams: 1) a monocultural team with high performance, 2) a monocultural team with low performance, 3) a multicultural team with high performance, and 4) a multicultural team with low performance (see section 3.2 for full explanation). The teams were compared based on the variable that changed, so:

1. The relationship between *observed psychological safety* in *monocultural vs multicultural teams* was obtained by comparing team 1 and team 3, and by comparing team 2 and team 4.

2. The relationship between *observed psychological safety* and *individual job performance* was obtained by comparing team 1 and team 2, and by comparing team 3 and team 4.

Comparative frequency analysis A comparative frequency analysis was carried out by performing a *t*-test on the data. This *t*-test was used to compare the means between the two teams according to the comparisons described above. A *t*-test can only be done when the data was randomized (✓), the teams were independent (✓), and the data was nearly normally distributed (De Veaux et al., 2015). To test normality, A Shapiro-Wilk test was performed on the data (Thode, 2002). An unpaired Student *t*-test was carried out when the data in both teams were nearly normal and had the same variance. When the data was nearly normal in both teams, but the variances were not equal, an unpaired Welch *t*-test was performed. Whenever the data was not nearly normal, a Mann-Whitney *U*-test was carried out to be able to compare the means (Fernandez, 2020).

Correlation analysis To test whether there might be a correlation between perceived and observed psychological safety, a Spearman's Rank Order Correlation analysis (Akoglu, 2018) was performed in R-studio. This correlation is supported by Spearman's rho because the psychological safety survey is based on a Likert Scale, which is considered a continuous variable (Lubke & Muthén, 2004). The Likert Scale values are correlated to the frequencies of psychological safety behaviors. The rho can take numbers from -1 to 1, which establishes how negative or positive the relationship (Akoglu, 2018) between perceived and observed psychological safety might be.

3.5.3 Qualitative analysis of episodes

Lastly, an episode analysis was done to compare interesting episodes from the meetings observed, to be able to give some deeper insights about the psychological safety behaviors than the insights given in the frequency analysis. Episodes can be defined as important moments in a team's ongoing activities, where 'a team member characterized these episodes as occasions of heavy engagement, salient, interaction dynamics, and strategically important decisions' (Jarrett & Liu, 2016, p.370). Because this thesis explores the relationship between psychological safety and individual job performance, the episode analysis served to observe the psychological safety behavior of outstanding individuals in the individual job performance ratings. The outstanding participants were from the same teams that were compared to each other in the frequency analysis. This episode analysis provided an additional set of results to give answer to the research question, with special regard to the relationship between psychological safety and individual job performance. To be able to compare the two selected episodes and related participants, the focus was on the defensive and unsupportive behaviors. These behaviors were likely to be less common than the behaviors that positively influence psychological safety, so a better distinction between the levels of psychologically unsafe behavior of the participants could be made.

4. RESULTS

4.1 Qualitative analysis of videos

In this section, the results of the qualitative analysis of the videos per team are explained. These explanations are based on two tables, which are presented on page 7: Table 2, in which the standardized frequencies per observed behavior are reported, and Table 3, in which the standardized duration per observed behavior is indicated. This way, the percentages per meeting can be compared.

4.1.1 Frequency and Duration of Behaviors in Team 1

Team 1 was a monocultural, high performing team. When looking at the frequency of behaviors, 94.21% of the behaviors were positively influencing the psychological safety during the

meeting. Oppositely, 2.79% of all behaviors were negatively influencing the psychologically safe environment. The rest, 3%, could be characterized as Neutral Behaviors, which did not influence psychological safety in a negative or positive way.

When looking at the duration of the behaviors, the behaviors that positively influence psychological safety were seen in 85.80% of all behaviors observed. The negatively influencing behaviors of psychological safety were seen in 2.44% of all behaviors. The rest, 11.74%, could be characterized as Neutral Behaviors.

A total frequency of 1467 behaviors were observed in a meeting that lasted 57:45:41 (min, sec, decimals). The behaviors are an accumulation of behaviors performed by 8 individuals. Since this meeting lasted the longest, and had the most team members, this meeting also had the highest frequency and total duration of behaviors.

4.1.2 Frequency and Duration of Behaviors in Team 2

Team 2 was a monocultural, low performing team. The frequency of behaviors could be divided into 92.56% of behaviors that influenced psychological safety positively, and 7.44% of behaviors that influenced psychological safety negatively during this meeting. The other 1.07% were Neutral Behaviors, which did not have any positive or negative influence on the psychologically safe environment.

The accumulated duration of behaviors is also shown in the table. In this meeting, 74.38% of the full duration of behaviors positively influenced psychological safety. Contrarily, 12.28% of the behaviors' duration had a negative influence on psychological safety. The remaining 13.33% of behaviors when looking at duration were considered neutral.

The total frequency of observed psychological safety behaviors was 1210, in a meeting of 53:03:92 (min, sec, decimals). All behaviors in this table are a total of behaviors performed by 5 team members, since not all team members were present during this meeting.

4.1.3 Frequency and Duration of Behaviors in Team 3

Team 3 was a multicultural, high performing team. As can be seen in Table 2, the behaviors that have a positive impact on psychological safety are observed in 91.44% of all behaviors seen. On the other hand, 7.19% of all behaviors observed have negatively impacted the psychologically safe environment. The remaining 1.37% observed have been Neutral Behaviors.

When looking at the accumulated duration of all behaviors observed in Table 3, 79.02% of the time behaviors were shown have positively influenced psychological safety. Defensive Voice Behaviors, Silence Behaviors, Defensive Silence Behaviors, and Unsupportive Behaviors were shown 16.12% of the time, which negatively influenced psychological safety. Silence Behaviors were shown most in this team when it comes to duration, which is interesting to note. Lastly, 1.37% of behaviors was considered neutral.

The duration of the meeting was 34:59:16 (min, sec, decimals), which explains why the total frequency, 584, and accumulated duration, 7921.540 seconds, of behaviors in this meeting were the lowest. The meeting was held by 6 team members.

4.1.4 Frequency and Duration of Behaviors in Team 4

Team 4 was a multicultural, low performing team. A total frequency of behaviors of 914 was observed, under which 91.94% positively influenced psychological safety in the meeting. The negatively influencing behaviors on psychological safety were seen 7.85% of the total frequency of observed behaviors. The rest of the behaviors observed were Neutral Behaviors, which resulted in a percentage of 0.22%.

Of the accumulated duration of behaviors, 88.98% of the time the behaviors influenced psychological safety positively during the meeting. Oppositely, 10.8% of the time the behaviors observed were negatively influencing the psychological safety. The remaining 0.21% of the accumulated duration were Neutral Behaviors.

The meeting had a length of 57:36:29 (min, sec, decimals), and was held by 6 team members. Therefore, the total accumulated duration of 14274.077 seconds consisted of all durations of behavior of 6 people.

4.1.5 Comparison: Observed Psychological Safety in Monocultural vs Multicultural Teams

Team 1 and Team 3 can be compared to investigate the relationship between observed psychological safety in mono- vs multicultural teams. Voice Behaviors, Silence Behaviors, Defensive Voice Behaviors, Defensive Silence Behaviors, Collaboration Behaviors (but, with a much longer duration in Team 1), and Unsupportive Behaviors are the behaviors that are more frequently seen in the multicultural, high performing team (Team 3). This indicates that all the negatively influencing behaviors on psychological safety are seen more frequently in the meeting of Team 3 than in the meeting of Team 1. Learning or Improvement Oriented Behaviors and Familiarity Behaviors have higher frequencies in Team 1, which is the monocultural, high performing team.

Furthermore, Team 2 and Team 4 can also be compared to evaluate the possible relationship between observed psychological safety and cultural diversity. In the multicultural, low performing team (Team 4), Silence Behaviors, Defensive Silence Behaviors, Collaboration Behaviors, and Unsupportive Behaviors are more frequently observed than in the monocultural, low performing team (Team 2). On the other hand, Voice Behaviors, Defensive Voice Behaviors, Learning or Improvement Behaviors, and Familiarity Behaviors are seen more frequently in Team 2. Interestingly, Team 4 has a higher duration in Voice Behaviors, and Team 2 has a higher duration in Silence Behaviors, which contradicts the frequencies. Here, more psychologically unsafe behaviors were seen in Team 4 than in Team 2.

To conclude, especially psychologically unsafe behaviors were more frequently observed in the multicultural teams. In addition, Learning or Improvement Oriented Behaviors and Familiarity Behaviors were more common in the monocultural teams.

4.1.6 Comparison: Observed Psychological Safety vs Individual Job Performance

Team 1 and Team 2 can be compared to evaluate the relationship between observed psychological safety and individual job performance. The low performing monocultural team (Team 2) has more Voice Behaviors, Defensive Voice Behaviors, Defensive Silence Behaviors, Unsupportive Behaviors, Learning or Improvement Oriented Behaviors, and Familiarity Behaviors. When looking at duration, there has been much more Silence Behavior as well in Team 2. From this can be concluded that all psychologically unsafe behaviors have been observed more in Team 2 than in Team 1. However, Familiarity Behaviors are observed more in the meeting of Team 2, too, especially when looking at the duration of Familiarity Behaviors. Collaboration Behaviors are observed more in Team 1, which was the high performing monocultural team.

Additionally, Team 3 and Team 4 can be compared to investigate the possible relationship between observed psychological safety and job performance. Voice Behaviors, Defensive Voice Behaviors, Defensive Silence Behaviors, Unsupportive Behaviors, Learning or Improvement Oriented Behaviors and Familiarity Behaviors are observed more frequently in Team 4 than in Team 3. This establishes that the psychologically unsafe behaviors are observed more in the team that is low performing, which was also the case when comparing Team 1 and Team 2. However, Silence Behaviors are observed more in Team 3, the high performing multicultural team, especially when looking at duration. Collaboration Behaviors are seen more frequently in Team 3, but have a higher duration in Team 4.

From these results, it can be concluded that especially psychologically unsafe behaviors were more frequently observed in the meetings of teams with a low average rating of individual job performance. This difference was larger than when comparing the mono- vs multicultural teams. However, Voice Behaviors, Familiarity Behaviors and Learning or Improvement Oriented Behaviors were also observed more in the low performing teams. Lastly, Collaboration Behaviors were more common in the high performing teams.

Table 2: Standardized frequencies per psychological safety behavior (maximum frequency per behavior is underlined)

	Team 1	Team 2	Team 3	Team 4
Voice Behaviors	13.77%	19.17%	14.73%	17.43%
Defensive Voice Behaviors	0.14%	3.47%	0.68%	2.40%
Silence Behaviors	2.04%	1.74%	4.28%	2.18%
Defensive Silence Behaviors	0.61%	1.90%	1.37%	2.40%
Collaboration Behaviors	67.42%	56.61%	69.18%	65.03%
Unsupportive Behaviors	0.00%	0.33%	0.86%	0.87%
Learning or Improvement Oriented Behaviors	5.93%	6.28%	4.62%	5.12%
Familiarity Behaviors	7.09%	10.50%	2.91%	4.36%
Neutral Behaviors	3.00%	1.07%	1.37%	0.22%

Table 3: Standardized duration per psychological safety behavior (maximum duration per behavior is underlined)

	Team 1	Team 2	Team 3	Team 4
Voice Behaviors	11.47%	15.04%	20.15%	18.20%
Defensive Voice Behaviors	0.01%	2.30%	0.47%	2.24%
Silence Behaviors	1.90%	8.62%	15.07%	6.55%
Defensive Silence Behaviors	0.53%	1.29%	0.42%	1.81%
Collaboration Behaviors	69.48%	44.99%	55.25%	67.41%
Unsupportive Behaviors	0.00%	0.07%	0.16%	0.20%
Learning or Improvement Oriented Behaviors	2.43%	3.15%	2.60%	2.12%
Familiarity Behaviors	2.42%	11.20%	1.02%	1.25%
Neutral Behaviors	11.74%	13.33%	4.86%	0.21%

4.2 Exploratory Statistical Comparative Frequency Analysis

This section is divided into two parts: 1) a description of the comparative frequency analysis used to explain the relationship between psychological safety in mono- vs multicultural teams, and 2) a description of the comparative frequency analysis used to explain the relationship between psychological safety and job performance. For these two parts, Appendices A to D are used to base the explanation of results on. These appendices show what assumptions were used to indicate what *t*-test needed to be performed, including some descriptive statistics to be able to draw conclusions. The *t*-tests were performed per behavior, to be able to draw detailed conclusions in the comparisons.

4.2.1 Exploratory Statistical Comparison: Observed Psychological Safety in Monocultural vs Multicultural Teams

The relationship between *observed psychological safety* and *cultural diversity* was obtained by comparing Team 1 and Team 3, and by comparing Team 2 and Team 4. In both comparisons, Neutral Behaviors were significantly different in all teams, however, since this does not have any negative or positive impact on psychological safety, neutral behaviors are not further explored.

In Appendix A, the monocultural, high performing team (Team 1) and the multicultural, high performing team (Team 3) are being compared based on the frequencies of psychological safety behaviors. The frequencies that were significantly different are Collaboration Behaviors, Unsupportive Behaviors, and Familiarity Behaviors. The other behaviors did not have a *p*-value that was below the critical value of $\alpha = .05$, which means that those frequencies were not significantly different.

Appendix B describes the comparative frequency analysis between the monocultural, low performing team (Team 2) and

the multicultural, low performing team (Team 4). Here, only the frequencies of Familiarity Behaviors were seen as significantly different. The other frequencies of behaviors did not have a *p*-value that was low enough, which indicates that those frequencies did not differ significantly enough.

4.2.2 Exploratory Statistical Comparison: Observed Psychological Safety vs Individual Job Performance

The relationship between *observed psychological safety* and *individual job performance* was obtained by comparing Team 1 and Team 2, and by comparing Team 3 and Team 4. Once again, Neutral Behaviors were significantly different in these comparisons, however, they are not further elaborated on because of the same reason.

As can be seen in Appendix C, there was a significant difference between the high performing monocultural team (Team 1) and the low performing monocultural team (Team 2), when it came to the following behaviors: Voice Behaviors, Defensive Voice Behaviors, Defensive Silence Behaviors, Collaboration Behaviors, Unsupportive Behaviors, and Familiarity Behaviors. Silence Behaviors and Learning or Improvement Oriented Behaviors are not significantly different when the frequency of observed behaviors was compared during the two meetings.

In Appendix D, Teams 3 and 4 were compared in terms of the frequency of psychological safety behaviors. The outcomes of the *t*-tests indicated that when the high performing multicultural team (Team 3) was compared with the low performing multicultural team (Team 4), no frequency of behavior was significantly different from each other, except for the Neutral Behaviors. This means that there was no statistical evidence that the differences between observed psychological safety and individual job performance is not due to chance in this comparison.

4.3 Correlation analysis: Observed vs Perceived Psychological Safety

Since this exploratory research draws conclusions based on *observing* psychological safety, this innovative method might lead to different results when comparing it to using perceived psychological safety measures. To be able to correlate the perceived psychological safety values given by each team member with the observed psychological safety measures, the two variables needed to be monotonically related to each other. That is why only the correlation between the standardized frequencies of positive psychological safety behaviors was correlated with the values given by each team member. When individuals performed psychologically unsafe behavior, this mostly had impact on other participants and not themselves, which made the relationship between observed negative behaviors per participant and perceived individual psychological safety invalid.

When correlating the frequencies of observed positive psychological safety behaviors per individual with the individual grade for perceived psychological safety, a Spearman's ρ of .27 was obtained. However, since the test had a *p*-value of .20, this Spearman's ρ is not significant when setting a critical value of $\alpha = .05$. Therefore, there is no statistical evidence that the differences between perceived and observed psychological safety is not due to chance.

4.4 Qualitative analysis of episodes

Participant 5 of Team 4 was the only one that scored 7 points on all four items of the individual job performance survey. Contrarily, participant 7 of Team 2 scored the lowest with an average score of 3.75. Two episodes are taken where these two participants are having a conversation about how the sprint was going. The episodes taken are both from when the two participants share their frustrations during the sprint. These are interesting parts of the meeting, since here the psychologically unsafe behaviors are most likely to be shown (Defensive Voice Behaviors, Defensive Silence Behaviors, Unsupportive Behaviors). Because the same type of conversation is taken in the two episodes, the results can be well compared.

4.4.1 Episode of Participant with the highest rating of Job Performance

In Appendix E the episode can be found for the participant with the highest Job Performance rating, which is participant 5 from Team 4. This episode was taken from the Retrospective meeting, which lasts 50 seconds. Prior to this episode, the behavior of participant 1 was observed as Defensive Voice Behavior directed at participant 5, from which participant 5 became frustrated. The episode begins at the moment participant 5 expressed his/her frustration. This episode was mostly a conversation between participant 1 and participant 5, where both participants expressed their frustrations towards each other in an aggressive tone. The other team members were occasionally showing some collaboration behaviors in the form of actively making eye contact. At the beginning of the episode, participant 5 tried to explain in an aggressive tone that there had been a difference in what has been communicated to the two, since they were not present at the same times during stand-up meetings. Immediately after, participant 1 showed some Defensive Silence Behaviors in the form of frowning the eyebrows, sighing and shaking his/her head. These behaviors were interpreted as belittling, and very clearly showing his/her annoyance towards participant 5. Participant 5 continued with his/her point, also marked as a Defensive Voice Behavior, since this was again interpreted as voice in an aggressive tone. Here he/she also made use of aggressive body language by making large gestures during his/her speech, directed towards participant 1. Participant 1 continued in a defensive tone with another argument, which again sounded very belittling towards participant 5. After this argument was made, participant 5 calmed down, and started using Voice Behavior as his/her means of communication. Even though participant 1 stayed in his/her aggressive tone, participant 5 stayed calm, and tried to explain the situation further. In between the Defensive Voice Behaviors of participant 1, participant 5 showed a Defensive Silence Behavior one more time, by using quite some space for his arm movements. However, there had still been an interesting switch in the middle of the episode from a use of defensive (aggressive) language, to Voice Behaviors when looking at participant 5.

4.4.2 Episode of Participant with the lowest rating of Job Performance

Appendix F describes the episode for the participant with the lowest ratings for Job Performance, that is participant 7 of Team 2. This episode is taken from the Retrospective meeting, and lasts 71 seconds. Here, participant 7 complained about an external factor from the Dutch organization that had an impact on the outcomes of the team. He/she described the problem, during which he/she evidently showed his/her emotions. The emotion that is shown most clearly is annoyance. This participant took it off in an aggressive tone on the other team members, while the others were not able to do something about it as well. The other team members showed comprehension, by reacting with Voice Behaviors and Collaboration Behaviors. Here, the team members were making eye contact with participant 7, reacting to his/her frustrations calmly, and trying to find solutions by offering help. At one point, participant 4 was trying to make the situation a little less tense by making a few jokes. Even though this had a positive

impact on the rest of the team, since they showed Familiarity Behaviors and Collaboration Behaviors in the form of laughing and agreeing after that, participant 5 did not perform any positively influencing behaviors on psychological safety, other than actively making eye contact. Even though participant 6 offered to help with the situation towards the end of the episode, participant 7 still stayed in the negative tone, and ended the episode with another defensive statement in an aggressive tone.

4.4.3 Comparison between episodes

The difference between these two episodes that can be seen most clearly, is how participant 5 of Team 4 calms down during his/her frustration, even though another team member has not ended his/her Defensive Behaviors. On the other hand, participant 7 in Team 2 did not calm down and stayed in his/her aggressive tone, even though the other team members were calm and did not perform any psychologically unsafe behaviors. Another difference is that in Team 4, the Defensive Behaviors are aimed at the two individuals having the conversation. On the other hand, in the meeting of Team 2, the Defensive Behaviors observed of participant 7 were aimed at the whole team. This might have had more impact on the psychological safety of the whole team than in the meeting of Team 4.

5. DISCUSSION

5.1 Theoretical implications

5.1.1 Psychological Safety and Monocultural vs Multicultural Teams

Since the comparative frequency analysis indicated that only the Collaboration Behaviors and Unsupportive Behaviors were significantly different in the comparison between the monocultural, high performing team (Team 1) and the multicultural, high performing team (Team 3), and Familiarity Behaviors in both comparisons, the difference seen between mono- and multicultural teams is rather small. However, the direction of this difference aligns with the literature. More specifically, Unsupportive Behaviors were observed more in the multicultural teams, which might be due to the factor of embracement or rejection of cultural differences (Leersnyder et al., 2022). In addition, Farley et al. (2022) described how team members from an international context are often perceived as less competent than the local ones, which might explain the difference in Unsupportive Behaviors, too. The results of this thesis support the previous research (Farley et al., 2022) in that the local team members were observed to perform the Unsupportive Behaviors more frequently than the international team members in these multicultural teams.

Additionally, Collaboration Behaviors and Familiarity Behaviors were observed more in the monocultural teams. Familiarity Behaviors might be lower in multicultural teams because of the communication challenges related to different perspectives, expectations and language-related barriers established by Glikson et al. (2016). It was clearly seen that jokes were being made more easily in the Dutch meetings than when the main language was English. This aligns with research done by Bell and Attardo (2010), where data showed that nonnative speakers had a harder time with humor in a second language than native speakers. Collaboration Behaviors were observed more in monocultural teams, which can be linked to cultural differences (Hughes, 2008). Indeed, multiple papers have underlined differences between cultures in the perception of making active eye contact. For instance, compared to the Dutch or European culture, Latin American and Asian cultures might not perceive eye contact as a Collaboration Behavior, since they generally find it disrespectful or uncomfortable (Akechi et al., 2013; Schneider & Barsoux, 2003). Actively making eye contact was observed most in the Collaboration Behaviors, which might explain why Collaboration Behaviors were observed less in the multicultural teams.

5.1.2 Psychological Safety vs Individual Job Performance

The results for the relationship between the level of observed psychological safety and individual job performance has been quite strong. Only the comparative frequency analysis between the high performing multicultural team (Team 3) and the low performing multicultural team (Team 4) did not indicate significant differences between the means, but the results of the rest of the analyses were aligned with the previous literature. In particular, in the teams with a low job performance, the level of observed psychological safety was also low. This is supported by the findings of Miao et al. (2019) and Glikson et al. (2016). However, Cauwelier et al. (2016) also mentioned team learning as a factor for this positive relationship. This is an interesting variable, since together with the negatively influencing behaviors on psychological safety observed in the low performing teams, Voice Behaviors and Learning or Improvement Oriented Behaviors are also seen more frequently in the low performing teams. These behaviors included informing about issues or mistakes, asking questions, providing feedback and speaking up with ideas for improvement, all of which can be identified as behaviors that positively influence the development of team learning (Cauwelier et al., 2016). Since the high performing teams therefore did not show more team learning behaviors than the low performing teams, this finding did not align with the research of Cauwelier et al. (2016).

From the findings of the episode analysis can be concluded that the participant with the highest rating for individual job performance was better at keeping his/her own psychological safety level high than the participant with the lowest ratings. Furthermore, the participant with the lowest rating had bigger impact on the whole team with his/her psychologically unsafe behavior when compared to the person having the highest ratings for individual job performance. These results coincide with the literature of Miao et al. (2019) and Glikson et al. (2016), where a positive relationship was also established. Additionally, the results of this analysis align with the findings of Cauwelier et al. (2016), where team learning played a role in the positive relationship between psychological safety and job performance. This is because the participant with the highest rating for job performance strengthened team learning, by switching to Voice Behaviors; He/she started informing about issues and building arguments constructively. Since the participant with the lowest rating of individual job performance performed none of the team learning behaviors (i.e. ‘seeking for feedback’, ‘looking for help’, ‘speaking up about concerns or mistakes’, and ‘innovative behavior and boundary spanning’; Cauwelier et al., 2016, p. 462), this participant did not influence team learning positively, and therefore psychological safety and individual job performance might have been lower, too.

5.1.3 Observed Psychological Safety vs Perceived Psychological Safety

When looking at the correlation analysis between observed psychological safety and perceived psychological safety, there was no statistical evidence that there might be a correlation between observed- and perceived psychological safety in this research. This might be due to a few reasons. Firstly, the values for the perceived psychological safety could have been subject to self-report biases (Bauhoff, 2011), which might have resulted in the data being less reliable. Secondly, since the survey data consisted of only three items, the survey might have captured too little details to accurately measure perceived psychological

safety (O'Donovan et al., 2020). Thirdly, the observed psychological safety data has been based on interpretations by the coders. As the research progressed, it was very difficult for the coders to reach an acceptable interrater reliability, which is why the observed psychological safety data might have included some errors.

The reasons mentioned above might have affected the outcome of why it is not possible to say whether there is a correlation between perceived and observed psychological safety in this research. However, O'Donovan et al. (2020) pointed out that the results of the two measures were prone to differences, and that triangulating the results can lead to a better understanding of the differences and similarities.

5.2 Practical implications

When combining the previous findings, there are some practical implications suggested. Psychological safety has a positive effect on individual job performance. Since individual job performance is a very important factor for the survival of an organization (Shanafelt & Noseworthy, 2017) and therefore important to maximize, organizations and their managers might want to increase their levels of psychological safety in meetings. Agile (HR) managers can do this by raising awareness about psychological safety through increasing employee knowledge. This can be done by giving (mandatory) training to communicate what the effect is of negatively influencing behaviors on the psychologically safe work environment. When people are aware, they might be able to change their own behavior positively, or correct someone if they feel like their behavior is not beneficial for the psychological safety of the team.

Furthermore, the results related to the differences between mono- and multicultural teams highlighted that multicultural teams seemed to manifest less psychologically safe behaviors than monocultural teams. Because the relationship between mono- vs multicultural teams and observed psychological safety was rather weak, organizations do not need to change teams accordingly. However, raising awareness might improve psychological safety especially in the multicultural teams. Research has explored several reasons for why multicultural teams experience less psychological safety, among which knowledge about another culture and the embracement of different cultures can be influenced most by an organization. Hence, to increase knowledge and embracement of different cultures, organizations can organize get-togethers when a project team has just been made, or publish interviews of employees from different cultural backgrounds. They can also provide training to increase the knowledge and awareness about cultures employees might come into a project group with.

5.3 Limitations and further research

As all research, this research has a few limitations that need to be considered. Firstly, the sample consisted of agile teams that voluntarily participated in this study, which might have led to sample selection bias, meaning that the teams that participated were those performing higher than average. Although there were some teams that did seem to perform less well than others, future research might add valuable insights when the teams researched are more randomly selected.

Secondly, the sample size consisted of teams that were all taken from the same organization in the Netherlands. Even though the teams fit in the requirements for this thesis, generalizability of the results cannot be claimed. Future research might add to this research by analyzing teams across organizations, preferably across countries, too.

Thirdly, the codebook used for this research is still in development. Consequently, one code of behavior in the codebook used was sensitive to multiple interpretations. This was the code for Collaboration Behavior, where "Active Listening", in the form of actively making eye contact, was particularly hard to interpret. Even though this code made it difficult to reach an acceptable interrater reliability between the two independent coders, the other behaviors were coded almost the same between the two coders. Therefore, it is advised to change the description of the code for Active Listening in the codebook provided by the OBCC group for future research. According to Spataro and Bloch (2017), active listening can be defined as when a 'listener gives the speaker full attention via inquiry, reflection, respect, and empathy' (p. 1). When connecting this to actively making eye contact, it should therefore not be a problem when an object that is contributing to the conversation is shortly looked at, which is not yet a description of the code. Also, finishing each other's' sentences can be seen as active listening when considering this definition. These suggestions might help further research to increase the interrater reliability of the codes used for the results.

Furthermore, even though the sample purposively consisted of agile teams that had meetings in person - since body language, active listening, or disengagement are more easily observable -, research has indicated that there are differences in psychological safety when it comes to virtual team meetings (Lechner & Mortlock, 2021). Future research might add some interesting findings when the same set of psychological safety behaviors are explored in a virtual setting.

Lastly, future research might be able to add a comparative duration analysis to the methodology. Even though a comparative frequency analysis can already be used to draw relevant conclusions, as can be seen in the description of the qualitative analysis of videos, the durations of behaviors have different distributions in the meetings than the frequency of behaviors. When adding a comparative duration analysis, different results might bring other conclusions.

6. CONCLUSION

This exploratory research investigated whether there are possible differences in observed psychological safety between mono- and multicultural agile teams, and whether there is a relationship between observed psychological safety and individual job performance. The mixed-method research design obtained results through a qualitative analysis of recorded Retrospective meetings, comparative frequency analyses, a correlation analysis, and an episode analysis. When combining the findings, monocultural agile teams seem to have a slightly higher level of observed psychological safety in the meetings than multicultural agile teams. This is mostly due to a higher presence of psychologically unsafe behaviors in the multicultural agile teams than in monocultural agile teams. Furthermore, a positive relationship has been established between observed psychological safety and individual job performance. Practical implications from this thesis suggest that organizations should increase awareness for the effects of psychological safety on individual job performance, and increase cultural knowledge and embracement of cultures.

7. REFERENCES

Aghina, W., Handscomb, C., Ludolph, J., Rona, D. and West, D. (2020), *Enterprise Agility: Buzz or Business Impact?*, McKinsey & Company, 20 March, available at: https://www.mckinsey.com/business-functions/organization/our-insights/enterprise-agility-buzz-or-business-impact

Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013). Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative Ratings. *PLOS ONE*, *8*(3), e59312. https://doi.org/10.1371/journal.pone.0059312

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Battistella, C., Bortolotti, T., Boscari, S., Nonino, F. and Palombi, G. (2023). The impact of cultural dimensions on project management performance. *International Journal of Organizational Analysis*, Vol. ahead-of-print No. ahead-of-print. https://doi-org.ezproxy2.utwente.nl/10.1108/IJOA-11-2022-3498

Bauhoff, S. (2011). Systematic self-report bias in health data: impact on estimating cross-sectional and treatment effects. *Health Serv Outcomes Res Method*, 11, 44–53. https://doi-org.ezproxy2.utwente.nl/10.1007/s10742-011-0069-3

Bell, N., & Attardo, S. (2010). Failed humor: Issues in non-native speakers' appreciation and understanding of humor. *Intercultural Pragmatics*, *7*(3). https://doi.org/10.1515/iprg.2010.019

Bergmann, B., & Schaeppi, J. (2016). A data-driven approach to group creativity. Harvard Business Review, 12th July 2016.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bresó, I., Gracia, F. J., Latorre, F., & Peiró, J. M. (2008). Development and validation of the Team Learning Questionnaire. *Comportamento Organizacional e Gestão*, *14*(2), 145–160.

Cai, Z., Huang, Q., Liu, H., & Wang, X. (2018). Improving the agility of employees through enterprise social media: The mediating role of psychological conditions. *International Journal of Information Management*, *38*(1), 52–63. https://doi.org/10.1016/j.ijinfomgt.2017.09.001

Calnan, M., & Rozen, A. (2019). ING's Agile Transformation—Teaching an Elephant to Race. *Journal of creating value*, *5*(2), 190–209. https://doi.org/10.1177/2394964319875601

Cauwelier, P., Ribiere, V., & Bennet, A. (2016). Team psychological safety and team learning: a cultural perspective. *The Learning Organization*. https://doi.org/10.1108/tlo-05-2016-0029

Christopher, M. (2000). The Agile Supply Chain. *Industrial Marketing Management*, *29*(1), 37–44. https://doi.org/10.1016/s0019-8501(99)00110-8

Cooper, R. G., & Sommer, A. F. (2016). The Agile-Stage-Gate Hybrid Model: A Promising New Approach and a New Research Opportunity. *Journal of Product Innovation Management*, *33*(5), 513–526. https://doi.org/10.1111/jpim.12314

Creswell, J. W., Clark, V. P., & Garrett, A. L. (2003). Advanced mixed methods research designs. *Handbook of mixed methods in social and behavioural research. Thousand Oaks, CA: Sage*, 209-240

De Leersnyder, J., Gündemir, S., & Agirdag, O. (2021). Diversity approaches matter in international classrooms: how a multicultural approach buffers against cultural misunderstandings and encourages inclusion and psychological safety. *Studies in Higher Education*, *47*(9), 1903–1920. https://doi.org/10.1080/03075079.2021.1983534

De Veaux, R. D., Velleman, P., & Bock, D. E. (2016). *Stats: Data and Models, Global Edition* (4th ed.). Pearson Education Limited

Dibble, R., Henderson, L., & Burns, Z. C. (2019). The Impact of Students' Cultural Intelligence on Their Psychological Safety in Global Virtual Project Teams. *Journal of Teaching in International Business*, *30*(1), 33–56. https://doi.org/10.1080/08975930.2019.1627980

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information & Software Technology*, *50*(9–10), 833–859. https://doi.org/10.1016/j.infsof.2008.01.006

Earley, P., & Ang, S. (2004). Cultural intelligence: individual interactions across cultures. *Choice Reviews Online*, *41*(08), 41–4760. https://doi.org/10.5860/choice.41-4760

Edmondson, A. C. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Edmondson, A.C. (2004), "Psychological safety, trust, and learning: a group-level lens", in Kramer, R. and Cook, K. (Eds), *Trust and Distrust in Organizations: Dilemmas and Approaches*, Russel Sage Foundation, New York, NY.

Farley, S., Dawson, J., Greenaway, T., Meade, K., & Ibar, D. N. H. (2022). Does International Status Affect Competence Ratings in Newly Formed Multinational Teams? The Role of Psychological Safety Growth and Verbal Behavior. *Journal of Cross-Cultural Psychology*, *53*(10), 1277–1299. https://doi.org/10.1177/00220221221115654

Fernandez, J. (2020, April 11). *The statistical analysis t-test explained for beginners and experts*. Towards Data Science. Retrieved May 7, 2023, from https://towardsdatascience.com/the-statistical-analysis-t-test-explained-for-beginners-and-experts-fd0e358bbb62

Gibson, C. B., Cooper, C. D., & Conger, J. A. (2009). Do you see what we see? The complex effects of perceptual distance between leaders and teams. *Journal of Applied Psychology*, *94*(1), 62–76. https://doi.org/10.1037/a0013073

Glikson, E., Harush, R., Kim, Y., Woolley, A. W., & Erez, M. (2016). Psychological Safety and Collective Intelligence in Multicultural Globally Dispersed Teams. *Social Science Research Network*. https://doi.org/10.2139/ssrn.3813887

Gonzalez-Diaz, R., & Bustamante-Cabrera, G. I. (2021). Predictive Sequential Research Design to Study

Complex Social Phenomena. *Entropy*, *23*(5), 627. https://doi.org/10.3390/e23050627

Grapenthin, S., Poggel, S., Book, M., & Gruhn, V. (2015). Improving task breakdown comprehensiveness in agile projects with an Interaction Room. *Information & Software Technology*, *67*, 254–264. https://doi.org/10.1016/j.infsof.2015.07.008

Groves, K. S., & Feyerherm, A. (2011). Leader Cultural Intelligence in Context. *Group & Organization Management*, *36*(5), 535–566. https://doi.org/10.1177/1059601111415664

Hughes, R. (2008). *Patient Safety and Quality: An Evidence-based Handbook for Nurses*. Department of Health and Human Services.

Jarrett, M., & Liu, F. (2016). "Zooming With". *Organizational Research Methods*, *21*(2), 366–385. https://doi.org/10.1177/1094428116656238

Kim, Y., Engel, D. M., Woolley, A. W., Lin, J., McArthur, N., & Malone, T. W. (2017). What Makes a Strong Team? *Conference on Computer Supported Cooperative Work*. https://doi.org/10.1145/2998181.2998185

Lechner, A. J., & Mortlock, J. T. (2021). How to create psychological safety in virtual teams. *Organizational Dynamics*, *51*(2), 100849. https://doi.org/10.1016/j.orgdyn.2021.100849

Leifels, K., & Bowen, P. (2021). The dark side of teamwork–the relationship between social stressors, social resources and team member well-being in monocultural and multicultural work teams. *Cross cultural & strategic management*, *28*(4), 867–893. https://doi.org/10.1108/ccsm-08-2020-0172

Lindsjørn, Y., Sjøberg, D. I. K., Dingsøyr, T., Bergersen, G. R., & Dybå, T. (2016). Teamwork quality and project success in software development: A survey of agile development teams. *Journal of Systems and Software*, *122*, 274–286. https://doi.org/10.1016/j.jss.2016.09.028

Lubke, G. H., & Muthén, B. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling*, *11*(4), 514–534. https://doi.org/10.1207/s15328007sem1104_2

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/bm.2012.031

Mesmer-Magnus, J., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, *94*(2), 535–546. https://doi.org/10.1037/a0013773

Miao, Q., Eva, N., Newman, A., & Cooper, B. (2019). CEO Entrepreneurial Leadership and Performance Outcomes of Top Management Teams in Entrepreneurial Ventures: The Mediating Effects of Psychological Safety. *Journal of Small Business Management*, *57*(3), 1119–1135. https://doi.org/10.1111/jsbm.12465

Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team: A case study of a Scrum project. *Information & Software Technology*, *52*(5), 480–491. https://doi.org/10.1016/j.infsof.2009.11.004

Morrison, E. W. (2014). Employee Voice and Silence. *Annual review of organizational psychology and organizational behavior*, *1*(1), 173–197. https://doi.org/10.1146/annurev-orgpsych-031413-091328

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. https://doi.org/10.5465/amr.2000.3707697

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review*, *27*(3), 521–535. https://doi.org/10.1016/j.hrmr.2017.01.001

O'Donovan, R., Van Dun, D. H., & McAuliffe, E. (2020). Measuring psychological safety in healthcare teams: developing an observational measure to complement survey methods. *BMC Medical Research Methodology*, *20*(1). https://doi.org/10.1186/s12874-020-01066-z

Paasivaara, M. (2017). Adopting SAFe to Scale Agile in a Globally Distributed Organization. *International Conference on Global Software Engineering*. https://doi.org/10.1109/icgse.2017.15

Paasivaara, M., Durasiewicz, S., & Lassenius, C. (2009). Using Scrum in Distributed Agile Development: A Multiple Case Study. *International Conference on Global Software Engineering*. https://doi.org/10.1109/icgse.2009.27

Rattini, V. (2023). Worker autonomy and performance: Evidence from a real-effort experiment. *Journal of Economics and Management Strategy*. https://doi.org/10.1111/jems.12511

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Schneider, S. C., & Barsoux, J. (2003). *Managing Across Cultures* (2de editie). Pearson Eduction Limited. https://books.google.nl/books?hl=en&lr=&id=KGIX5wSNQmcC&oi=fnd&pg=PR11&dq=making+eye+contact+in+different+cultures&ots=mhywKHXpPC&sig=MhiqWkjYk_z_a78JrlQobHSp3uo&redir_esc=y#v=onepage&q=Eye%20contact&f=false

Serrador, P., & Pinto, J. K. (2015). Does Agile work? — A quantitative analysis of agile project success. *International Journal of Project Management*, *33*(5), 1040–1051. https://doi.org/10.1016/j.ijproman.2015.01.006

Shanafelt, T. D., & Noseworthy, J. H. (2017). Executive Leadership and Physician Well-being. *Mayo Clinic Proceedings*, *92*(1), 129–146. https://doi.org/10.1016/j.mayocp.2016.10.004

Šmite, D., Moe, N. B., Floryan, M., Gonzalez-Huerta, J., Dorner, M., & Sablis, A. (2023). Decentralized decision-making and scaled autonomy at Spotify. *Journal of Systems and Software*, *200*, 111649. https://doi.org/10.1016/j.jss.2023.111649

Sommer, A. F. (2019). Agile Transformation at LEGO Group. *Research-technology Management*, *62*(5), 20–29. https://doi.org/10.1080/08956308.2019.1638486

Spataro, S. E., & Bloch, J. (2017). "Can You Repeat That?" Teaching Active Listening in Management Education.

Journal of Management Education, *42*(2), 168–198. https://doi.org/10.1177/1052562917748696

Stahl, G. K., & Maznevski, M. L. (2021). Unraveling the effects of cultural diversity in teams: A retrospective of research on multicultural work groups and an agenda for future research. *Journal of International Business Studies*. https://doi.org/10.1057/s41267-020-00389-9

Stahl, G. K., Mäkelä, K., Zander, L., & Maznevski, M. L. (2010). A look at the bright side of multicultural team diversity. *Scandinavian Journal of Management*, *26*(4), 439–447. https://doi.org/10.1016/j.scaman.2010.09.009

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Thode, H. C. (2002). *Testing For Normality*. CRC Press.

Thorgren, S., & Caiman, E. (2019). The Role of Psychological Safety in Implementing Agile Methods across Cultures. *Research-technology Management*, *62*(2), 31–39. https://doi.org/10.1080/08956308.2019.1563436

Van Dyne, L., Ang, S., & Botero, I. C. (2003). Conceptualizing Employee Silence and Employee Voice as Multidimensional Constructs*. *Journal of Management Studies*, *40*(6), 1359–1392. https://doi.org/10.1111/1467-6486.00384

Zielske, M., & Held, T. (2022). Agile methods used by traditional logistics companies and logistics start-ups: a systematic literature review. *Journal of Systems and Software*, *190*, 111328. https://doi.org/10.1016/j.jss.2022.111328

8. APPENDIX

Appendix A: Comparative Frequency Analysis between Team 1 and Team 3, critical value α = .05

Behavior	Team	Assumption 1: Randomized data?	Assumption 2: Independent teams?	Assumption 3: Data nearly normally distributed?	Mean	Std. Deviation	Assumption 3.1: Equal variance?	*t*-test & *p*-value
Voice Behaviors	1 3	✓ ✓	✓	Yes Yes	0.017 0.025	0.006 0.010	Yes	Unpaired Student *t*-test *p* = .111
Defensive Voice Behaviors	1 3	✓ ✓	✓	No No	0.000 0.001	0.000 0.003	-	Mann Whitney *U*-test *p* = .928
Silence Behaviors	1 3	✓ ✓	✓	No Yes	0.003 0.007	0.003 0.008	-	Mann Whitney *U*-test *p* = .295
Defensive Silence Behaviors	1 3	✓ ✓	✓	Yes Yes	0.001 0.002	0.001 0.003	No	Unpaired Welch *t*-test *p* = .212
Collaboration Behaviors	1 3	✓ ✓	✓	Yes Yes	0.084 0.115	0.011 0.027	No	Unpaired Welch *t*-test *p* = .037
Unsupportive Behaviors	1 3	✓ ✓	✓	No Yes	0.000 0.001	0.000 0.001	-	Mann Whitney *U*-test *p* = .012
Learning or Improvement Oriented Behaviors	1 3	✓ ✓	✓	Yes Yes	0.007 0.008	0.006 0.005	Yes	Unpaired Student *t*-test *p* = .927
Familiarity Behaviors	1 3	✓ ✓	✓	Yes Yes	0.009 0.005	0.004 0.001	No	Unpaired Welch *t*-test *p* = .015
Neutral Behaviors	1 3	✓ ✓	✓	Yes No	0.004 0.002	0.001 0.001	-	Mann Whitney *U*-test *p* = .012

Appendix B: Comparative Frequency Analysis between Team 2 and Team 4, critical value $\alpha = .05$

Behavior	Team	Assumption 1: Randomized data?	Assumption 2: Independent teams?	Assumption 3: Data nearly normally distributed?	Mean	Std. Deviation	Assumption 3.1: Equal variance?	*t*-test & *p*-value
Voice Behaviors	2 4	✓ ✓	✓	Yes Yes	0.038 0.029	0.007 0.017	Yes	Unpaired Student *t*-test $p = .304$
Defensive Voice Behaviors	2 4	✓ ✓	✓	Yes No	0.007 0.004	0.007 0.008	-	Mann-Whitney *U*-test $p = .261$
Silence Behaviors	2 4	✓ ✓	✓	Yes No	0.003 0.004	0.003 0.008	-	Mann-Whitney *U*-test $p = .453$
Defensive Silence Behaviors	2 4	✓ ✓	✓	Yes No	0.004 0.004	0.001 0.007	-	Mann-Whitney *U*-test $p = .081$
Collaboration Behaviors	2 4	✓ ✓	✓	No Yes	0.112 0.108	0.014 0.026	-	Mann-Whitney *U*-test $p = .647$
Unsupportive Behaviors	2 4	✓ ✓	✓	Yes Yes	0.001 0.001	0.001 0.002	Yes	Unpaired Student *t*-test $p = .339$
Learning or Improvement Oriented Behaviors	2 4	✓ ✓	✓	Yes Yes	0.012 0.009	0.008 0.007	Yes	Unpaired Student *t*-test $p = .413$
Familiarity Behaviors	2 4	✓ ✓	✓	Yes Yes	0.021 0.007	0.004 0.004	Yes	Unpaired Student *t*-test $p < .001$
Neutral Behaviors	2 4	✓ ✓	✓	No No	0.002 0.000	0.001 0.001	-	Mann-Whitney *U*-test $p = .026$

Appendix C: Comparative Frequency Analysis Between Team 1 and Team 2, critical value α = .05

Behavior	Team	Assumption 1: Randomized data?	Assumption 2: Independent teams?	Assumption 3: Data nearly normally distributed?	Mean	Std. Deviation	Assumption 3.1: Equal variance?	*t*-test & *p*-value
Voice Behaviors	1 2	✓ ✓	✓	Yes Yes	0.017 0.038	0.006 0.007	Yes	Unpaired Student *t*-test $p < .001$
Defensive Voice Behaviors	1 2	✓ ✓	✓	No Yes	0.0002 0.007	0.0003 0.007	-	Mann-Whitney *U*-test $p = .021$
Silence Behaviors	1 2	✓ ✓	✓	No Yes	0.003 0.003	0.003 0.003	-	Mann-Whitney *U*-test $p = .456$
Defensive Silence Behaviors	1 2	✓ ✓	✓	Yes Yes	0.001 0.004	0.001 0.001	Yes	Unpaired Student *t*-test $p < .001$
Collaboration Behaviors	1 2	✓ ✓	✓	Yes No	0.084 0.112	0.011 0.014	-	Mann-Whitney *U*-test $p = .011$
Unsupportive Behaviors	1 2	✓ ✓	✓	No Yes	0.000 0.001	0.000 0.001	-	Mann-Whitney *U*-test $p = .022$
Learning or Improvement Oriented Behaviors	1 2	✓ ✓	✓	Yes Yes	0.007 0.012	0.006 0.008	Yes	Unpaired Student *t*-test $p = .23$
Familiarity Behaviors	1 2	✓ ✓	✓	Yes Yes	0.009 0.020	0.004 0.004	Yes	Unpaired Student *t*-test $p < .001$
Neutral Behaviors	1 2	✓ ✓	✓	Yes No	0.004 0.002	0.001 0.001	-	Mann-Whitney *U*-test $p = .014$

Appendix D: Comparative Frequency Analysis between Team 3 and Team 4, critical value $\alpha = .05$

Behavior	Team	Assumption 1: Randomized data?	Assumption 2: Independent teams?	Assumption 3: Data nearly normally distributed?	Mean	Std. Deviation	Assumption 3.1: Equal variance?	*t*-test & *p*-value
Voice Behaviors	3 4	✓ ✓	✓	Yes Yes	0.025 0.029	0.010 0.017	Yes	Unpaired Student *t*-test $p = .581$
Defensive Voice Behaviors	3 4	✓ ✓	✓	No No	0.001 0.004	0.003 0.008	-	Mann-Whitney *U*-test $p = .341$
Silence Behaviors	3 4	✓ ✓	✓	Yes No	0.007 0.004	0.008 0.008	-	Mann-Whitney *U*-test $p = .142$
Defensive Silence Behaviors	3 4	✓ ✓	✓	Yes No	0.002 0.004	0.003 0.007	-	Mann-Whitney *U*-test $p = 1$
Collaboration Behaviors	3 4	✓ ✓	✓	Yes Yes	0.115 0.108	0.027 0.026	Yes	Unpaired Student *t*-test $p = .664$
Unsupportive Behaviors	3 4	✓ ✓	✓	Yes Yes	0.001 0.001	0.001 0.002	Yes	Unpaired Student *t*-test $p = .977$
Learning or Improvement Oriented Behaviors	3 4	✓ ✓	✓	Yes Yes	0.008 0.009	0.005 0.007	Yes	Unpaired Student *t*-test $p = .821$
Familiarity Behaviors	3 4	✓ ✓	✓	Yes Yes	0.005 0.007	0.001 0.004	No	Unpaired Welch *t*-test $p = .215$
Neutral Behaviors	3 4	✓ ✓	✓	No No	0.002 0.000	0.001 0.001	-	Mann-Whitney *U*-test $p = .022$

Appendix E: Episode Participant with highest Job Performance Rating

Meeting 07001: 20.27 min. to 21.17 min.

Transcript	Participant	Behavior
F5: <> I talked about this in the stand-up but you missed it. That was at ten O'clock and you were not there, and we did another one, and I was not there. And when I came back we did not do a model stand-up, the second stand-up, but would be the third, because at ten o'clock I did one.	F5	Defensive Voice Behavior
	F1	Collaboration Behavior
	F7	Collaboration Behavior
	F6	Collaboration Behavior
	F1	Defensive Silence Behavior
F1: pff ok.	F1	Collaboration Behavior
	F1	(continued) Defensive Silence Behavior
F5: so if you don't know about this, with that that was blocked, you were not in this stand up.	F5	Defensive Voice Behavior
	F5	Defensive Silence Behavior
	F1	Collaboration Behavior
	F4	Collaboration Behavior
	F5	Collaboration Behavior
F1: ok, so now were are going to also be very nitty gritty-	F1	Defensive Voice Behavior
F5: no	F5	Voice Behavior
F1: if certain people are not in the stand-up, but I don't – I don't	F1	(continued) Defensive Voice Behavior
F5: I just put also that – one way to solve that, that they have to solve that, is to put the guys from <> -	F5	(continued) Voice Behavior
F1: know what is going on and why you are so being so frustrated.	F4	Collaboration Behavior
	F5	Defensive Silence Behavior
	F1	(continued) Defensive Voice Behavior
F5: <> who work with me, I just did this.	F1	Collaboration Behavior
	F5	(continued) Voice Behavior
F1: I know	F1	Collaboration Behavior
	F6	Collaboration Behavior
	F1	Defensive Silence Behavior
F5: without this, I cannot follow that.	F5	(continued) Voice Behavior

Appendix F: Episode Participant with lowest Job Performance Rating

Meeting 08001: 27.12 min. to 28.23 min.

Transcript	Participant	Behavior
F5: steeds nieuwe portals.		
F7 Ja.	F7	Collaboration Behavior
	F5	Collaboration Behavior
F4: we hebben er weer een.	F4	Familiarity Behavior
F7: we hebben er weer een nieuwe bij.	F7	Voice Behavior
<>. Dus ik word er een beetje- beetje hopeloos van uh. <> die lopen- die veranderen elke week zo een beetje, of om de week. En dan hoor je bij het nieuwe portaal af, wat half of niet ingericht is. Waar je nog eerst achteraan moet, voordat je weer met je proces verder kan. Ja en ik word daar een beetje moe van.	F7	Defensive Voice Behavior
	F4	Collaboration Behavior
	F5	Collaboration Behavior
	F6	Silence Behavior
F5: ja. Nee uh-	F5	Collaboration Behavior
	F3	Collaboration Behavior
F7: dus daarom zei ik vanochtend ook van-		
F4: Dat kost dagen	F7	Defensive Voice Behavior
	F4	Voice Behavior
F7: ik zei ook tegen hem- op een gegeven moment tegen hem <naam> nou zo van weet je, ga totaal geen handleidingen meer schrijven, want zodra ik de ene handleiding geschreven heb, krijgen we een nieuwe <paal>, en dan moet ik dat weer aan ga zitten passen.	F7	(continued) Defensive Voice Behavior
	F6	Collaboration Behavior
	F5	Collaboration Behavior
	F4	Collaboration Behavior
F4: zet je iemand anders op je overdracht?	F4	Familiarity Behavior
F6: ja, goed	F6	(continued) Collaboration Behavior
	F4	(continued) Familiarity Behavior
F4: nou, hier heb je hem.	F6	Familiarity Behavior
F7: nou <>	F7	Voice Behavior
F6: nou weet je, ik ga hem hier- ik neem hem mee naar <naam>, misschien is het goed om daarna even te overleggen met <naam> specifiek hierover te hebben.	F6	Voice Behavior
	F5	Collaboration Behavior
	F4	Collaboration Behavior
F4: nou, ik vind het- ik vind het heel bijzonder dat wij alleen tegen dit soort dingen aanlopen.	F4	Voice Behavior
	F7	Collaboration Behavior
	F6	Collaboration Behavior
	F3	Collaboration Behavior
	F5	Collaboration Behavior
	F7	Collaboration Behavior
F6: ja, ik vind het ook heel raar. Maar-	F6	Voice Behavior
F4: volgens mij is de rest totaal niet met beheer bezig van dit soort dingen.	F4	(continued) Voice Behavior
F7: Ja	F7	Collaboration Behavior
F3: dat idee heb ik ook een beetje.	F3	(continued) Collaboration Behavior

	F4	Collaboration Behavior
	F7	Collaboration Behavior
F6: nou, B zestien is die uh <>-	F6	Voice Behavior
F7: ja, het enige wat er wordt gezegd is dit zouden overige types ook moeten doen. Goh, misschien handig.	F7	Defensive Voice Behavior