Measuring and Affecting Psychological Safety in Software Teams: A Case Study

1st Mikkel Agerlin Christensen
IT University of Copenhagen
Copenhagen, Denmark
magc@itu.dk

ABSTRACT

This paper presents a replication study used to measure levels of psychological safety in teams in the field of medicine, adapted and implemented in the field of software engineering. Replicated survey and observation methods were applied to measure levels of psychological safety within two teams from a Danish software company. After the first round of measurements, an intervention workshop on spreading awareness of psychological safety was held, before a second round of measurements was taken.

Results indicate that this intervention affected psychological safety in the teams, as both teams moved just under one standard deviation between the pre- and post-workshop measurements, with one team increasing its score, and the other decreasing. It is suspected that these movements indicate that increasing awareness improved the identification of psychological safety in team events, causing increases in identification of both positive and negative behaviours.

Index Terms—**Psychological safety, software, teams, team collaboration, software engineering, survey, observation, case study**

I. INTRODUCTION

In 1999, Amy Edmondson published her seminal work on psychological safety, laying the foundation for research on the subject [1]. She stated that psychological safety covers many aspects of human interaction, such as "a belief that one will not be punished or humiliated for speaking up with ideas, questions, concerns, or mistakes" [2].

Edmondson found that psychological safety existed in most interpersonal interactions, and that psychological safety was a key component in team learning and innovation. Fifteen years later Google found psychological safety to be the number one predictor of team effectiveness [3]. Psychological safety is hard to measure. While attempts have been made in the medical field [4], research on psychological safety in the domain of software is sparse [5].

For this reason, this study has looked into the field of medicine, and chosen to replicate the work of O'Donovan et al. in [6], using surveys and observations to measure the psychological safety of teams, in a software context. This will be done by applying O'Donovan et al.'s method in a Danish software company.

O'Donovan et al.'s study measured existing levels of psychological safety within the teams, intervened on the aware-ness of psychological safety, and finally measured the effect of this intervention on the psychological safety of the team.

This study aims to investigate the application of O'Donovan et al.'s method in a software engineering context. Additionally, it aims to measure whether spreading awareness of the psychological safety construct can act as an intervention on it, affecting change in a company's psychological safety levels - hopefully for the better. It will do so based on two research questions:

RQ1: To what extent can the measurements of O'Donovan et al. be used to measure psychological safety in a software team?

RQ2: Can increasing awareness of the psychological safety construct affect the psychological safety of a software team?

Below, the academic background and related work of the study are presented.

II. BACKGROUND

This section will outline the academic background of the psychological safety construct, while the related work section presents other empirical studies on psychological safety conducted in both medicine and software industries.

Amy Edmondson coined the term psychological safety in 1999, defining it as "a shared belief held by members of a team that the team is safe for interpersonal risk taking" [1]. Edmondson argues that these risks plays a role in almost all interpersonal interaction in work teams, and that the team upholds a shared belief of the level of psychological safety in the team, depending on the prevalence of those risks. This shared belief in, and the accompanying level of, psychological safety is associated with positive learning behaviour, which "mediates between team psychological safety and team performance" [1]. Most people in organizations are constantly evaluated, either by peers or superiors, consciously or subconsciously [7]. Edmondson argues that this evaluation comes with an implicit concern for one's personal and professional image, and with it, four primary interpersonal risks for learning behaviour; worrying about being seen as either ignorant, incompetent, negative or disruptive [7]. Fear of being seen as ignorant might prevent an employee from asking a question that they think they should know the answer to, while a fear of being seen as incompetent might cause reluctance in willingness to experiment, or reluctance to admit crucial mistakes. A fear

of being seen as negative often prevents critical feedback, while a fear of being disruptive might prevent employees from daring to challenge the status quo. A psychologically safe work environment is one that is conducive to taking these risks [7]. In these work environments, employees engage in a "cognitive group-level construct" [7], wherein participants assess interpersonal risks of their actions (such as the actions mentioned above), based on their perception of how this action will be perceived. This means that some actions that might be performed (and deemed acceptable) in psychologically safe work environments can be entirely impossible to take in a different environment.

Taking on such interpersonal risks could be viewed as a matter of trust within the team. Edmondson clarified that although psychological safety shares many commonalities with trust, they are not the same [7]. She points out the following three differences:

1) **Temporal Immediacy**: When evaluating the trade-offs of taking the above mentioned risks, short term interpersonal consequences are in focus. Edmondson gives an example of "a nurse facing the decision of whether to ask a physician in the unit about a medication dosage she suspects is erroneous may be so focused on the potential immediate consequences of asking this question, such as being scolded and humiliated for being ignorant, that she temporarily discounts the longer-term consequence of not speaking up — that is, the harm that may be caused to a patient." [7]

2) **Focus on "self" versus "other"**: Edmondson argues that "Trust involves giving others the benefit of the doubt", while psychological safety is an expression of "whether others will give you the benefit of the doubt when, for instance, you have made a mistake or asked an apparently stupid question." [7]. This results in an internal focus that is "in contrast to the focus on others' future actions implicit in the construct of trust" [7].

3) **Levels of analysis**: Edmondson argues that psychological safety is a group-level construct, shared among members of the team as "the way things are around here" [7]. In contrast, "trust pertains primarily to a dyadic relationship, even if that dyad is sometimes conceptualized as consisting of large entities." [7].

III. RELATED WORK

Recently, Google's study "Aristotle" [3] found psychological safety to be the number one predictor of team effectiveness. Additionally, Google's 2019 State of Dev-Ops named "Culture of Psychological Safety" a major contributor to "... organizational performance, and productivity, showing that growing and fostering a healthy culture reaps benefits for organizations and individuals." [8], a result found independently through the application of two separate research models. This indicates that although research on psychological safety is lacking in the field of software engineering, it has been identified as something of great importance to software teams.

Several related works have been conducted within the field of medicine, however. In particular, research has been done by O'Donovan et al. on both exploring [6] , measuring and intervening [4] on psychological safety, in addition to a systematic literature review [9] of psychological safety literature within the field of medicine. The work of O'Donovan et al. in [4] aimed to develop a method to to measure levels of psychological safety in teams, which has been replicated in this study. This method of data-gathering was designed specifically to inform interventions on psychological safety. It takes a triangulation approach, combining surveys and observations for a more in-depth view on measurements of psychological safety. This method was chosen after identifying situations wherein measurements of psychological safety from surveys alone were found to not accurately represent the levels of psychological safety within the surveyed teams.

A presentation of the individual elements of survey, observation and, composite measurement methods presented in [4], and replicated in this study, can be found in the next Section.

IV. METHOD

The empirical work conducted in this study was carried out in a Danish software company, which develops apps and websites for medical purposes in Denmark. It consisted of a replication of the survey and observation methods described in [4]. These methods were conducted on two participating teams. For the purposes of this study, they will be labeled simply as Team 1 and Team 2.

Early in the project, both survey and observation methods were tested with a different team from the company, in order to practice and verify the methods in a software context. When early testing had been conducted, measuring and affecting the psychological safety of Team 1 and Team 2 began.

The research into those two teams consisted of the following events:

1) Sending out surveys to participating teams in order to Measure the levels of psychological safety of each team

2) Recording stand ups of each team for observation analysis

3) Composing and analysing survey and observation data into a final measurement of psychological safety

4) Intervening on psychological safety through a workshop with the teams

5) Repeat of 1-3 after the intervention workshop, using the same methods, in order to the analyze effects of intervention

Below, these elements of empirical work are discussed. First, the survey, observation and data analysis methods are presented. Second, method adaptations are discussed. Third, the design and contents of the awareness workshop are presented. Finally, threats to the validity of this method are presented.

A. Survey

The survey method replicates the survey presented in [4], attached in Appendix A. Their survey design draws on the

work of several previous authors, such as Van Dyne et al. [10], Le Pine and Van Dyne [11], and Amy Edmondson [7], whose 7-question survey [1] was adopted into the survey. The survey questions were answered on a Likert scale of 1-7, with one being strongly disagree, and 7 being strongly agree. The survey aims to investigate individual evaluations of psychological safety in various work place situations, such as holding back feedback or questions, or surveying the degree to which the respondents feel comfortable admitting to a mistake. Surveys were distributed to the two participating teams twice, one set before the intervention and one after. The second round of surveys was identical except for a single question asking respondents whether they had taken the survey from the first round (pre-intervention). Any respondent answering negatively to this, was excluded from results analysis, as their answer could not be used to measure differences in pre- and post-workshop levels of psychological safety, when they did not answer the pre-intervention survey.

The survey was intended to be taken shortly after a team event, in order to catch respondents in a moment of clear memory of team interactions. The survey included three questions concerning the degree to which the meeting, the respondent's, and the team's behaviour differed from other similar meetings. These questions were included in order to identify situations in which survey results stemmed from reactions to an unusual event.

For the full survey results, see Section V.

B. Observation

The observation procedure replicates the observation method described in [4], attached in Appendix B. The observation method presents a framework of 31 observable behaviours within 7 categories; voice, defensive voice, silence behaviours, supportive, unsupportive, learning or improvement-oriented and familiarity type behaviours [4]. These are in part inspired by the work of Van Dyne et al. in [10], which conceptualizes the constructs of employee silence and voice behaviours. The observation method was intended to give an outside observer evaluation of levels of psychological safety, and act as triangulation method to the surveys, in order to evaluate the impacts of self-reporting, if survey answers reported something widely different from the observed behaviours.

Eight stand up meetings of Team 1 and Team 2 were recorded for observation. These sets of recordings were made in two different time frames; one set before the intervention workshop, and and one set recorded after. For each team, a pre- and post-workshop recording was chosen for analysis, picking those that were most similar in length. These recordings were later analysed in detail to identify instances of the behaviours from the framework. This was done using the tool ChronoViz, which allows time-stamping events in the recordings on a timeline. An example of this can be seen in Figure 1.

Using this tool, multiple reviews of the recorded stand up meetings were conducted, each time focusing on identifying subsets of the 31 observed behaviours, in order to ensure that proper attention was paid to each behaviour. The observed
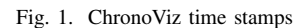

Fig. 1. ChronoViz time stamps

Fig. 2. Composite Measure Integrating Survey and Observation Components [4]

teams varied in size, and the recorded meetings varied in length, affecting the total number of identified behaviours in each meeting. However, upon selecting recordings for observation analysis, recordings of most similar lengths were chosen in order to minimize the effect of different meeting lengths on observation data. Due to time constraints of this study, 4 such observations were chosen for in-depth analysis, one for each team in the pre- and post-intervention phase respectively. Additionally, results were analysed in relation to the balance of supportive and unsupportive behaviours within each team, rather than a relative comparison of the raw number of supportive or unsupportive behaviours between teams, which further diminishes the effect of these differences on the analysis. Results from these observations can be seen in Section V.

C. Data analysis: Composing Survey and Observation Results

Once survey and observation data was gathered, results were compared for an overview of levels of psychological safety within the teams. This allowed for triangulation of findings, comparing psychological safety levels in observed behaviours to those from the self-reported surveys. An example of this with dummy data can be seen in Figure 2 (example taken from the original authors in [4]).

Two independent instances of this analysis was be conducted for each team, one for the measurements taken before

the intervention workshop, and one after. These analyses are presented as the results of this study (see Section V), and will be used to reason about the levels of psychological safety within the teams before and after the intervention workshop.

D. Method Adaptations to the Case

In applying the method developed in the field of medicine, several adaptations were made to apply it in in the field of software engineering. Additionally, some adaptations were made for the specific case, to work with the processes of distributed, agile software teams. The list of adaptations can be seen below.

The Team Leader: The replicated survey and observation methods differentiate heavily between interactions of team members, and those between team members and team leaders, scoring the team leader's interactions individually. In an agile software team, team leadership plays a very different role. Agile software teams believe that "The best architectures, requirements, and designs emerge from self-organizing teams." [12], commonly resulting in flatter hierarchies. The participating agile teams did not have a designated "team leader". The CTO reported that the participating teams referred directly to them as team leader. While this definition of team leader sufficed for the survey questions directed at team leadership, it required further adaptations in the observed events. While the CTO participated in these events, their role was not one of team member, and they remained largely passive throughout the events. Their main interaction was in setting up and closing down the meeting. While several observation metrics are identified in silent or passive behaviour (see Appendix B for the full list), these are all negative behaviours, such as closed body language, or showing signs of fear or disengagement. While these metrics are relevant, in analysing the observed events, they very rarely occurred, leaving the team leader section of the observation data almost empty. Having very few observation data points on the team leader meant that a single recorded behaviour could greatly skew the analysis outcome. For these reasons, the differentiation between team leader and team member interactions has not been included in the observation analysis. Instead, only the interaction of team members has been recorded. Had the CTO acted differently, and engaged in more of the observation metric behaviours (negative or otherwise), this choice would have been re-evaluated. The differentiation between psychological safety towards team members and the team leader is still present in the survey.

Patient Safety: The observation included observing behaviour related to patient safety. Since software companies do not generally interact with patients, this was instead changed to observing behaviours of "client safety". Client safety was defined as conversation based on concern for the clients' well-being or wishes, such as discussing whether a certain feature fulfills the critical needs of a client, or whether a feature exposes client data in a way that is not ethical.

Digital Communication: The replicated observation method was designed for in-person observations of employees work-

ing in co-located setups. Several elements of the observed behaviours are challenging to identify in a digital setting. The observation in this study was conducted on recordings of distributed team events (stand-ups) held over Slack video chat. Even though the meetings were carried out with participants using webcams, not all webcams were visible at all times in a call with many participants, due to Slack's design. Silence behaviours are more difficult to identify when participants are sitting behind a screen, as it can be more challenging to identify disengagement or closed body language in smaller video feeds. In a co-located setting, it is quite easy to determine if meeting participants are looking at, and engaging with, the speaker. In a digital setting, it isn't immediately clear what the participants are a looking at, especially if they have multiple monitors. Additionally, body language in front of a computer is inherently different from real world interaction, as most of the participants' movements relate to interacting with their digital equipment. Disengagement in the digital setting was primarily identified when participants looked away, or their eyes started scrolling sideways (indicating reading), or when participants started typing, in a context where it was not prompted for by the meeting, in addition to the familiar indicators from co-located settings (body language, facial expressions), whenever they were available.

E. The Awareness Workshop

In researching psychological safety, no known and proven mechanisms for affecting psychological safety were found. In fact, a systematic literature study showed that attempts at affecting psychological safety yielded inconsistent results [9]. This study therefore adopted an explorative approach to finding an alternative way to affect psychological safety levels in the teams.

In a meeting between the case company's CTO and the researchers, options for improving psychological safety in the teams were discussed. It was agreed that it would be a good approach to aim for team members understanding the topic, and to establish a common term (namely, psychological safety) among team members to describe the elements of everyday work life that psychological safety affects. The choice of focusing on awareness was primarily based on the meeting participants' own experiences of becoming aware of Edmondson's idea of psychological safety, which had enabled reflection about their own experiences in the work place (past or present), aiding them in reasoning about it and discussing it with colleagues. It was therefore decided in this meeting that spreading awareness of Edmondson's concept of psychological safety among the teams would be the target of the workshop, aiming to improve levels of psychological safety in the team through enabling similar reflection and discussion between team members.

Attempting to affect change of psychological safety levels is trying to implement behavioural change within the teams. Starting with targeting awareness follows the ADKAR model for change management, which describes awareness as the first step to enabling change [13]. Hiatt states that "[awareness]

is achieved when a person is aware of and understands the nature of the change, why it is needed, and the risks of not changing" [13]. The nature of the change as well as why the change was needed was a key part of the design of the awareness workshop. Additionally, several activities of the awareness workshop specifically involved exploring situations in which psychological safety was not present, and the risks and challenges that those situations brought with them.

The workshop consisted of the following events:

1) **Presentation**: An introductory presentation of what Psychological Safety is, and why a workshop was held on it.
2) **Ted Talk by Amy Edmondson**: Shared viewing of Amy Edmondson's TEDx Talk on psychological safety [2].
3) **Break-out room discussion**: In groups of three, participants were asked to recall and share previous work experiences in which they held back on opinions or questions. Each group then wrote down notes about their conversation in shared slide deck.
4) **Group Presentations**: Each group presented the notes they put on their slides during their break-out conversations.
5) **Applying the contents of the workshop to the work place**: Several initiatives to apply the contents and discussion that occured during the workshop were presented.
6) **Open floor feedback through MentiMeter**: The tool MentiMeter was used for anonymous open floor feedback in relation to the workshop and its contents.

It was decided that Amy Edmondson's talk should lay the foundation of the workshop and its following activities. As Edmondson is at the forefront of Psychological Safety research, and commonly performs presentations on it, we saw no reason to try and replicate her work, and instead shared the source directly.

The workshop was run digitally through Slack. It was followed by a a second round of surveys and observations, using the exact same method as the first round of surveys and observations. Herein, it would be visible whether raising awareness resulted in different levels of psychological safety in the surveys and observations. In spreading awareness and surveying again, it should be noted that the second survey measure was not intended to investigate the participants' awareness or understanding of what psychological safety is. The interest of spreading awareness was to see whether awareness of the construct of psychological safety alone could affect the survey respondents' answers, producing a measurable difference between the pre-workshop and the post-workshop levels of psychological safety within the teams.

F. Threats to Validity

1) **Case generalizability**: As this is a single-case study, results can not necessarily be generalized and applied to other companies directly, as findings could be dependant on factors unique to the case.

2) **Team Structure**: The company studied has an organizational structure that challenges formal definitions of teams and team leadership. Several team members floated between teams, and no team leader was defined outside of C-level positions having a natural leadership role. While this threat has been covered in part by the method adaptations mentioned above, it is possible that a more re-structural adaptation could be made, in order to account for a more fluid team structure in flat hierarchy.
3) **Existing degrees of psychological safety in company culture**: The company researched in this case study is one with a high culture of openness and safety. While the company had not worked with psychological safety prior to the study, many other similar concepts of employee well-being and efficiency had been explored. All teams surveyed (including the team from the testing round) scored high on levels of psychological safety, even before the workshop. The degree to which this affected the findings of this study will be discussed in Section VI.
4) **Time Frame and behavioural change**: As mentioned earlier, attempting to affect levels of psychological safety is trying to affect behavioural change. The study ran for 6 months, wherein only the latter half was spent measuring, affecting, and re-measuring levels of psychological safety within the teams. The degree to which any fundamental change in levels of psychological safety within the participating teams can happen within such a time frame is uncertain.
5) **Self-reporting**: Using surveys opens the study to self-reporting threats, such as participants misunderstanding questions, or mis-representing the way they broadly feel, based on how they feel in the moment of taking the survey. This is especially challenging in a survey of a more emotive nature.
6) **Observed events**: While the replicated observation method does not speak specifically as to what kind of meetings should be observed, stand up meetings carry a specific structure that affects the contents and forms of interaction during the meeting. Stand ups are usually short informative meetings rather than longer discussions, although discussion is encouraged when challenges are identified. As such, it is to be expected that stand ups are more inviting to some of the 31 observations behaviours, such as providing information, than others. However, as the outcome of the observation metric is a comparison between supportive and unsupportive behaviours, it does not matter whether the observed behaviours were spread over many categories, or a few select ones. It is possible that using a different event for observation, such a sprint retrospective, would have provided more varied data for analysis.
7) **Observing digitally versus in person**: While adaptations were made to observe in a digital setting rather than a co-located setting, a digital setting is arguably more limiting. This threat was accentuated by Slack

sometimes hiding participants' videos, as well as having rather small camera feeds for participants who were not currently speaking. It was not always possible to evaluate what a meeting participant was focused on, or whether they paid attention to the speaker. This is only a threat to observation of passive behaviours, however, as those who were speaking were put in enlarged camera feeds. Multiple reviews of the recorded observations were conducted to palliate this threat.

V. RESULTS

This section will present results of the data gathered through the survey and observation methods described in section IV. Results were gathered from two teams. Team 1 had 9 team members, while Team 2 had 8 team members. Outside of these, three developers aided both teams, as described in Section IV. It was not known which team they identified with in the survey. For a discussion of team structure and team membership, see Section IV-F.

Survey questions were answered on a scale of 1-7, with one being strongly disagree, and 7 being strongly agree, while observation scores are shown simply as counts of observed supportive and unsupportive behaviours. For a full breakdown of all identified behaviours in observations of the team meetings see Appendix C.

Below, results of each team from data gathering before and after the intervention workshop are presented.

A. Pre-workshop results

Below, results from data gathered before the intervention workshop are presented. In the pre-workshop survey, 5 out of 8 team team members responded from Team 1, while 9 responses was given to Team's 2 survey. Team 2 only had 7 team members fully allocated to their team, indicating that at least two of the three loosely coupled developers answered identifying as Team 2. All but a single team member from team 2 reported that the meeting which the survey was taken after was either not at all different, or neutral in comparison to similar meetings, across all three questions evaluating the meeting. The single team member from team 2 reported that the meeting was slightly different across all three questions.

The score of each team's pre-workshop survey can be seen in Table I.

TABLE I
SURVEY RESULTS

Team	Psychological Safety Towards Team Leader		Psychological Safety Towards other Team Members		Psychological Safety in relation to team as a whole	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Team 1	6.667	0.977	6.22857	1.215	6.8	0.414
Team 2	5.578	1.293	5.347	1.48	5.571	1.4

The count of observed behaviours identified in the pre-workshop events can be seen in Table II.

TABLE II
OBSERVATION RESULTS

Total observed behaviours		
Team 1 (8:30 minute meeting)		**Distribution**
categories indicating high PS	56	93.33%
categories indicating lower PS	4	6.67%
Team 2 (10 minute meeting)		
categories indicating high PS	44	75.86%
categories indicating lower PS	14	24.14%

B. Post-workshop results

Below, results from data gathered after the intervention workshop are presented. Table cells are colored green if a cell value was greater than in the pre-workshop results and red if it was lesser. In the post-workshop survey, 6 people from Team 1 participated, while Team 2 had 7 participants. The post-workshop survey had a question added, asking participants whether they responded to the first survey. All 6 respondents from Team 1 answered positively to this question, despite only 5 respondents from Team 1 answering the pre-workshop survey. As the surveys were conducted anonymously, it was not possible to identify whether a person had responded twice, or if someone had responded to the second survey, but not the first one, and misreported this. For this reason, all responses were kept for analysis purposes, since identification was necessary for elimination of incorrect data. A single respondent from Team 2 responded "no" to having been surveyed before. Their answer was not included in the results data.

The meeting that the survey was taken after was deemed neutral to not different, in comparison to similar meetings, by the vast majority of respondents, with no respondents claiming it to be very different.

The score of each team's post-workshop survey can be seen in Table III.

TABLE III
SURVEY RESULTS

Team	Psychological Safety Towards Team Leader		Psychological Safety Towards other Team Members		Psychological Safety in relation to team as a whole	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Team 1	6	1.274	6.024	1.615	6.667	1.188
Team 2	6.296	1.127	6.119	1.214	6.222	1.166

The count of observed behaviours identified in the post-workshop events can be seen in Table IV.

As is indicated in the full overview of observed behaviours in Appendix C, all of the unsupportive behaviours identified in both post- and pre-intervention observations were indications of disengagement or closed body language. No aggression, interruption or cold reactions were observed.

TABLE IV
OBSERVATION RESULTS

Total observed behaviours		
Team 1 (7 minute meeting)		**Distribution**
Categories indicating high PS	27	72.97%
categories indicating lower PS	10	27.03%
Team 2 (9:30 minute meeting)		
Categories indicating high PS	24	80.00%
Categories indicating lower PS	6	20.00%

VI. CONCLUSION AND DISCUSSION

This paper has outlined a survey and observation-based study replicating methods from the field of medicine on measuring and intervening on psychological safety in two teams at a Danish software company. Two measures were taken, with an intervention workshop on spreading awareness of psychological safety being conducted in between. The survey measure successfully captured nuanced levels of psychological safety within the teams, and was successfully applied to allow for measurements of pre- and post-intervention levels of psychological safety. The observation measure supported this with observation results for triangulation. However, the observation measure was limited by the lack of a defined team leader in the observed teams, which the observation method based on.

Results show that both teams measured highly in psychological safety in the pre-workshop survey, with Team 1 reporting higher than Team 2 in all three surveyed categories.

After holding an intervention workshop and re-measuring, Team 1's score decreased by just under one standard deviation, while Team 2's score increased by a similar amount, bringing the two teams' measurements of psychological safety closer together. This is consistent with observation data indicating the same moment for both teams. This indicates that the intervention workshop successfully affected levels of psychological in the teams, and that spreading awareness of the psychological safety construct can affect the levels of psychological safety within a team.

Awareness being able to cause both an increase and a decrease in levels of psychological safety could indicate that increasing awareness of psychological safety increased ability to identify the role of psychological safety in the teams' events, in both positive and negative situations. As such, it could be argued that both teams were brought closer to the "real" levels of psychological safety within the teams. This aligns with the fact that the teams' scores were moved closer together after the intervention. While psychological safety can differ from team to team, it seems reasonable to assume that multiple teams within a company might be closer together in levels of psychological safety, than teams from different companies, since they are part of the same larger group in the form of the company, which also shares a group level of psychological safety. Additionally, making employees aware of a potential point of conflict, as psychological safety can be, can result in the identification of more instances of such conflicts, causing measurements of levels of psychological safety to drop.

While these findings indicate movement, the movement is one measured between two measurement points. It is not possible to claim with certainty that this movement was directly caused by the intervention. Further research is needed to identify whether the identified movements predict a continuous trend, and whether this trend is one of moving teams closer towards a "real" measure, or if the two teams will continue to respectively increase and decrease their levels of psychological safety over time. For future work, a continuous, iterative approach of measurement, intervention, and re-measurement could be taken to identify such a trend. For conducting such work, the method applied could additionally be adapted further to better support a flatter organizational structure.

Overall, measurements of psychological safety seemed to accurately reflect the impression of an engaging and safe environment that was gotten from observing the teams' events, wherein team members interacted constructively and made jokes. The intervention workshop was an engaging event wherein several team members actively contributed to the conversation. Subsequent observed events further displayed this culture, which, while potentially affected by the intervention, remained one of high psychological safety.

REFERENCES

[1] A. Edmondson, "Psychological safety and learning behavior in work teams," *Administrative science quarterly*, vol. 44, no. 2, pp. 350–383, 1999.

[2] ——, "Building a psychologically safe workplace," 2014. [Online]. Available: https://www.youtube.com/watch?v=LhoLuui9gX8

[3] A. Dubey, "Project aristotle - understanding team effectiveness," 2014. [Online]. Available: https://rework.withgoogle.com/print/guides/5721312655835136/

[4] R. O'Donovan, D. Van Dun, and E. McAuliffe, "Measuring psychological safety in healthcare teams: developing an observational measure to complement survey methods," *BMC medical research methodology*, vol. 20, no. 1, pp. 1–17, 2020.

[5] P. Lenberg and R. Feldt, "Psychological safety and norm clarity in software engineering teams," in *Proceedings of the 11th international workshop on cooperative and human aspects of software engineering*, 2018, pp. 79–86.

[6] R. O'Donovan and E. McAuliffe, "Exploring psychological safety in healthcare teams to inform the development of interventions: combining observational, survey and interview data," *BMC Health Services Research*, vol. 20, no. 1, pp. 1–16, 2020.

[7] A. C. Edmondson, *Managing the risk of learning: Psychological safety in work teams*. Citeseer, 2002.

[8] N. Forsgren, D. Smith, J. Humble, and J. Frazelle, "2019 accelerate state of devops report," 2019.

[9] R. O'Donovan and E. McAuliffe, "A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour," *BMC health services research*, vol. 20, no. 1, pp. 1–11, 2020.

[10] L. V. Dyne, S. Ang, and I. C. Botero, "Conceptualizing employee silence and employee voice as multidimensional constructs," *Journal of management studies*, vol. 40, no. 6, pp. 1359–1392, 2003.

[11] J. A. LePine and L. Van Dyne, "Predicting voice behavior in work groups." *Journal of applied psychology*, vol. 83, no. 6, p. 853, 1998.

[12] A. Manifesto, "Manifesto for agile software development," 2001.

[13] J. Hiatt, *ADKAR: a model for change in business, government, and our community*. Prosci, 2006.

Table 5 Final Survey Items

Length of time working with this team:
Please respond to the following questions by indicating your response between 1 = strongly disagree and 10 = strongly agree

Section 1. Please answer the following questions in relation to your **team leader**

Questions	[b]1 Strongly disagree	2	3	4	5	6	7 Strongly agree
1. If I had a question or was unsure of something in relation to my role at work, I could ask my team leader							
2. I can communicate my opinions about work issues with my team leader							
3. I can speak up about personal problems or disagreements to my team leader							
4. I can speak up with recommendations/ideas for new projects or changes in procedures to my team leader							
5. If I made a mistake on this team, I would feel safe speaking up to my team leader							
6. If I saw a colleague making a mistake, I would feel safe speaking up to my team leader							
7. If I speak up/voice my opinion, I know that my input is valued by my team leader							
8. My team leader encourages and supports me to take on new tasks or to learn how to do things I have never done before.							
9. If I had a problem in this company, I could depend on my team leader to be my advocate							

Section 2. Please answer the following questions in relation to your **peers/the other members of your team**

Questions	[b]1 Strongly disagree	2	3	4	5	6	7 Strongly agree
10. If I had a question or was unsure of something in relation to my role at work, I could ask my peers							
11. I can communicate my opinions about work issues with my peers							
12. I can speak up about personal issues to my peers							
13. I can speak up with recommendations/ideas for new projects or changes in procedures to my peers							
14. If I made a mistake on this team, I would feel safe speaking up to my peers							
15. If I saw a colleague making a mistake, I would feel safe speaking up to this colleague							
16. If I speak up/voice my opinion, I know that my input is valued by my peers							

Section 3. Please answer in relation to your **team as a whole**

Questions	[b]1 Strongly disagree	2	3	4	5	6	7 Strongly agree
17. *It is easy to ask other members of this team for help[a]*							
18. People keep each other informed about work-related issues in the team							
19. There are real attempts to share information throughout the team							

*Compared to similar meetings with your team how **different** was:[b]*

Questions	Very different	Different	Slightly Different	Neutral	Slightly different	Not different	Not at all Different
20. *This meeting?*							
21. *Your behaviour during this meeting*							
22. *The behaviour of your colleagues*							

[a]Reverse scored
[b]The sections of the survey marked in italics were changed based on pilot tests

APPENDIX B
[4] - OBSERVATION

Observation Measure Used During Pilot Tests

Behaviours	**Team member** behaviour count*	**Team leader** behaviour count*
Voice Behaviours		
Communicating opinions to others even if they disagree		
Asking questions		
Providing information		
Providing feedback		
Providing help or solutions		
*Correcting others** *		
Defensive Voice Behaviours		
Denying faults or blame others		
Showing aggression (Raising voice, large gestures)		
Evading confrontation by focusing only on positives		
Silence Behaviours		
Facial expression indicates fear		
Facial expression indicates disengagement		
Closed body language (arms closed, lean backwards)		

*No eye contact (with speaker)		
Supportive Behaviours		
Sharing procedures, knowledge and experience		
*Sharing future plans** *		
Active listening (verify, paraphrase)		
Use of inclusive language such as "we"		
Agreeing/Responding positively or enthusiastically to input		
*Acknowledging achievements/ congratulating one another** *		
*Delegating tasks** *		
Unsupportive Behaviours		
Interrupting		
Discussions within small sub-groups		
Reacting cold/ignoring a joke		
Learning or Improvement Oriented Behaviours		
Reviewing own progress and performance		
Asking for feedback		

Asking for help or solutions		
Asking for input from all meeting participants		
*Informing the team about issues or mistakes related to patient safety**￼*		
Looking for improvement opportunities and speaking up with ideas		
Acknowledging own mistake		
Familiarity Behaviours		
Talking about personal, non-work matters (with team members)		
Talking about personal, non-work matters (with team leader)		
Laughing about a joke		
~~*Safety Oriented Behaviour~~		
~~*Leaders' words and deeds align~~		
~~*Informing the team about issues or mistakes related to patient safety~~		

Observer ratings after meeting: Rated from 1 (strongly disagree) to 7 (Strongly agree)

Observations	1 (strongly disagree)	2	3	4	5	6	7 Strongly agree

There was enough opportunity for participants to ask for help							
There was enough opportunity for participants to speak up							
There was enough opportunity for participants to discuss with the team leader							
Certain team members dominated the discussion							
Decisions were made together, by the entire team							
The atmosphere in this team was constructive							
People seemed genuine and not to hold back anything							

Meeting duration:

*These items have been crossed out as they were removed or moved to another category following the first three pilot tests.

**The items marked in italics were added following the first eight pilot tests.

Definition of Behaviours in Observation Measure

Behaviours	Definitions
Voice Behaviours	

Communicating opinions to others even if they disagree	Sharing opinions that contradict what others have said or that other team members disagree with.
Asking questions	Speaking up with any question.
Providing information	Providing others with relevant information, updating them on issues.
Providing feedback	(Constructively) Commenting on other's input or behaviour
Providing help or solutions	Offering to take part in correcting a problem or issue or suggesting a solution.
Correcting others	Speaking up to correct what is being said by another team member.
Defensive Voice Behaviours	
Denying faults or blame others	When discussing mistakes or problems, speaker denies own fault/blames others.
Showing aggression	Displays of aggression in voice or body language, i.e. raised voice, or large gestures.
Evading confrontation by focusing only on positives	When a difficult or confrontational issue is being discussed, the team member shifts focus to more positive issues rather than directly addressing the difficult issue.
Silence Behaviours	
Facial expression or body language indicates fear	Any indication of fear (i.e. darting eye movements or nervous shifting in chair) either when speaking, during silence or while listening to others.
Facial expression or body language indicates disengagement	Any indication of disengagement (i.e. looking around the room, at phone or body facing away from others) either when speaking, during silence or while listening to others.
Closed body language	Team members arms are closed, they are leaning backwards.

Supportive Behaviours	
Sharing procedures, knowledge and experience	Sharing a solutions or learning experiences contributing to teams understanding of the issue at hand and team learning.
Sharing future plans	Providing information on actions or procedures which will take place after meeting.
Active listening	Responding to what others are saying by paraphrasing or verifying what was said.
Use of inclusive language such as "we"	Speaker refers to the team using "we" rather than "I" and "you".
Agreeing/Responding positively or enthusiastically to input	Positive response to others i.e. agreeing, nodding or reinforcing what has been said.
Acknowledging achievements/ congratulating	Team members opening acknowledge achievements and congratulate one another on them.
Delegating tasks	Tasks or responsibilities are shared across the team.
Unsupportive Behaviours	
Interrupting	Team member talks over another or steers the discussion away to another topic.
Discussions within small sub-groups	A small number of team members have a sub-conversation during which they are temporarily disengaged form the team discussion and do not share with others.
Reacting cold/ignoring a joke	When a joke is made, the team member either does not laugh or continues as if they did not hear the joke.
Learning or Improvement Oriented Behaviours	
Reviewing own progress and performance	Discussions on reviewing initiatives or aspects of performance that are working well or need to be improved.

Asking for feedback	Team member requests for input from other team members on what they have said.
Asking for help or solutions	Speaking up with requests for help or suggestion of solutions from other team members.
Asking for input from all meeting participants	An open request to all team members for input.
Informing the team about issues or mistakes related to patient safety or staff safety	Speaking up about and discussing issues that are directly relevant to patient or staff safety.
Looking for improvement opportunities and speaking up with ideas	Speaking up with suggestions for improvement in relation to issue of relevance to the team.
Acknowledging own mistake	Team member acknowledges a mistake they made. The team member does not blame others or circumstances and acknowledges own short coming.
Familiarity Behaviours	
Talking about personal, non-work matters	Any discussion which is not about work
Laughing about a joke	Team laughs together at a joke or funny comment

TABLE V
BREAKDOWN OF OBSERVED BEHAVIOURS

Behaviour	Team 1 - Pre	Team 2 - Pre	Team 1 - Post	Team 2 - Post
Voice Behaviours				
Communicating opinions to others even if they disagree'	0	0	0	1
Asking questions'	7	1	0	5
Providing information'	14	12	9	7
Providing feedback'	0	0	1	0
Providing help or solutions'	1	0	0	0
Correcting others'	0	0	0	0
Defensive Voice Behaviours				
Denying faults or blame others'	0	0	0	0
Showing aggression (Raising voice large gestures)	0	0	0	0
Evading confrontation by focusing only on positives	0	0	0	0
Silence Behaviours				
Facial expression indicates fear'	0	0	0	0
Facial expression indicates disengagement'	0	4	6	0
Closed body language (arms closed lean backwards)'	1	0	0	3
Supportive Behaviours				
Sharing procedures knowledge and experience	6	3	3	0
Sharing future plans'	3	8	3	3
Active listening (verify paraphrase)'	1	0	1	0
Use of inclusive language such as "we"'	7	7	8	2
Agreeing/Responding positively or enthusiastically to input'	8	0	2	1
Acknowledging achievements/ congratulating one another'	0	0	1	0
Delegating tasks'	0	2	0	0
Unsupportive Behaviours				
Interrupting'	0	0	0	0
Discussions within small sub-groups'	0	0	0	0
Reacting cold/ignoring a joke'	0	0	0	0
Learning or Improvement Oriented Behaviour				
Reviewing own progress and performance'	1	8	1	1
Asking for feedback'	1	0	0	0
Asking for help or solutions'	3	0	0	1
Asking for input from all meeting participants'	0	0	0	0
Informing the team about issues or mistakes related to patient safety'	0	0	0	0
Looking for improvement opportunities and speaking up with ideas'	0	1	1	0
Acknowledging own mistake'	0	4	0	1
Familiarity Behaviours				
Talking about personal non-work matters (with team members)'	0	5	0	0
Talking about personal non-work matters (with team leader)'	0	0	0	1
Laughing about a joke'	7	3	1	4