

# Building a Toolbox for Working with Psychological Safety in Agile Software Teams

Mikkel Agerlin Christensen<sup>1</sup> and Paolo Tell<sup>1</sup>

IT University of Copenhagen, Copenhagen, Denmark,  
magc@itu.dk – pate@itu.dk

**Abstract.** This paper presents the design of eight tools for working with psychological safety in agile software teams, which were designed in collaboration with industry practitioners using design science. The tools were adopted over a two-week period by four Danish industry software teams and evaluated through team interviews and surveys. Results show that the designed tools can be successfully adopted and integrated in the practices of a software team. Participating teams found the tool format valuable, as it allowed them (i) to engage in discussions they were not always capable of having, (ii) to find the right shared vocabulary to frame these discussions, and (iii) to provide them with needed prompts to let such discussion surface. Finally, teams unanimously reported interest in the continued use of the designed tools.

**Keywords:** Psychological safety, agile, teams, design science

## 1 Introduction

In 1999, Edmondson published her seminal work on psychological safety [6], defining the term as “a shared belief held by members of a team that the team is safe for interpersonal risk taking” and laying the foundation for future research on the subject. 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 most important predictor of team effectiveness [5]. Though psychological safety is important to understand and attend to, it is hard to measure, and even harder to improve. Research on measurements of psychological safety have been published in the medical domain [16], but research on affecting change of psychological safety is sparse, especially in the domain of software engineering [12].

Due to this sparsity, a previous 6-month case study in a Danish software company was conducted [1], replicating survey and observation methods used to measure levels of psychological safety in the medical domain [16]. “Triggered by an industry [...] need that can be addressed by developing an artifact” [17], the study included an initial exploration focused on the design of an intervention through a workshop to affect change in levels of psychological safety.

To reduce the gap identified in [1] and [5], among others, and improve the understanding of viable practices for working with psychological safety in software teams, this paper aims to further explore this topic through the use of

the methodology presented by Peffers et al. [17] for conducting design science research, hence, by creating, evolving, and evaluating artifacts (*tools*) to assist and enable teams to work with psychological safety. In this paper, the term *tool* refers to the tangible, descriptive representation of an intervention activity designed to affect change (intervene) on levels of psychological safety. When referring to such *tools*, the following italicized format will be used: *tool*.

This paper presents the design and production of a toolbox comprising eight such *tools*, which can be selected and adopted by teams wishing to incorporate working with psychological safety into their practice. Four software teams participated in evaluations after implementing a selection of *tools* over a two weeks period. Through the design and evaluation of these *tools*, this paper aims to answer the following research questions:

**RQ1:** How can interventions on psychological safety be designed as actionable *tools* which enable agile software teams to work with psychological safety as part of their practices?

**RQ2:** To what extent can *tools* aid agile software teams in working with psychological safety?

The remainder of this paper is structured as follows: Section 2 presents related work, while Section 3 presents the method used to develop and evaluate the *tools*. Section 4 presents the design and evolution of the *tools*. Finally, results are presented in Section 5, discussed in Section 6 and concluded on in Section 7.

## 2 Related Work

Google’s study “Aristotle” found psychological safety to be the number one predictor of team effectiveness across 180 international teams [5]. Additionally, Google’s 2019 “State of Dev-Ops” named a “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” [10], a result found independently through the application of two separate research models. Of the five key dynamics found to be significant (psychological safety, dependability, structure and clarity, meaning of work, impact of work) they found that “Psychological safety was far and away the most important of the five dynamics we found – it’s the underpinning of the other four” [5]. This indicates that, despite the lack of research on the application in the domain of software, its importance is well-established.

**Measuring Levels of Psychological Safety.** Several attempts have been made at quantifying psychological safety in the medical domain. In particular, research has been done by O’Donovan et al. on both measuring [16] and intervening [14] on psychological safety. The work of O’Donovan et al. in [16] developed a method to measure levels of psychological safety in teams, which was replicated in the pre-study [1]. This method of data gathering was designed specifically to inform interventions on psychological safety. This method was replicated in the pre-study [1], in which explorative work on measuring and affecting change on levels of psychological safety within software teams was conducted in a 6-month

project with two teams from a Danish software company. In [1], the survey and observation methods for measuring psychological safety from [14, 16] were applied within the software domain, in order to measure the effects of intervening on psychological safety.

**Intervening on Psychological Safety.** In the pre-study, the measurements of O’Donovan et al [16] were used to measure levels of psychological safety before and after an intervention workshop aiming to heighten the awareness of psychological safety within the participating teams. While the project’s explorative (and short) nature was only an initial step towards the improvement of psychological safety, several of the lessons learned motivated this paper. Specifically, the workshop showed that awareness alone could act as an intervention on psychological safety, something which became an early inspiration for the *tool* concept. The measurement techniques used, while applied successfully to the domain of software, were deemed more appropriate for continued measurements over longer periods of time, and as such will not be re-used in this paper, given its short and exploratory nature. O’Donovan et al. also analysed outcomes of interventions to improve psychological safety in [15]. Herein they concluded that the reviewed attempts on improving psychological safety had mixed results, in part identifying that “multifaceted interventions may allow future studies to further investigate the efficacy or effectiveness of these interventions.” [15]. The *tools* designed in this paper explore such multifaceted intervention, with the intent of investigating their effectiveness within software teams.

### 3 Method

This paper expands on preliminary work [1] and, by following design science research guidelines (i.e., Hevner et al. [11], Peffers et al. [17], and Wieringa [20]), aims to answer the research questions providing knowledge supporting the design of solutions in the form of artifacts to real-world either *construction* or *improvement* problems [3].

Figure 1 depicts the four cycles followed for artifact (*tool*) design and the mapping of steps of the design science process proposed by Peffers et al. [17]. These cycles and the related process steps (depicted as squares) are detailed in chronological order in the following. Process steps involving external input, such a workshop with participants, are marked with a triangle corner. Artifact versions from each Cycle are depicted by the rounded squares at the top of each Cycle, with the details of their evolution being presented in Section 4.

**Cycle 0.** Initiated in [1], and leading to an objective-centered entry point (i.e., “triggered by an industry or research need that can be addressed by developing an artifact“ [17]), the cycle led to: the identification of the main challenges, the analysis of the motivation to solve these, and the objectives of a potential solution. This cycle founded the research questions and the insight of using a designed artifact to solve the challenges, with motivation drawn from the identified gap in research, and the insights provided by Google in [5].

4 Agerlin Christensen and Tell

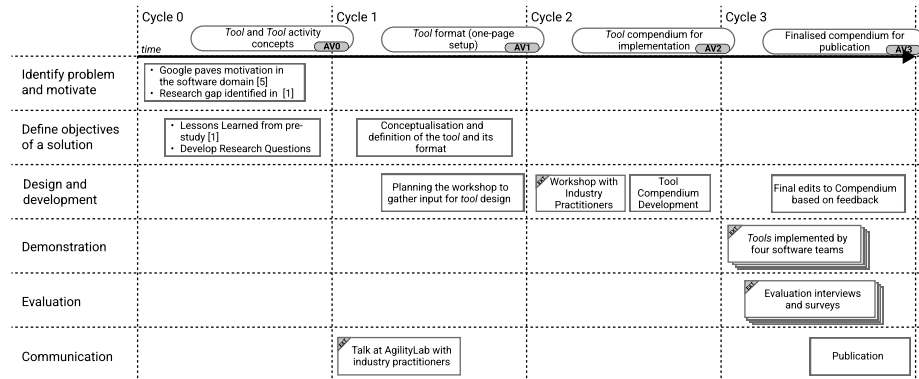


Fig. 1: Project activities based on the model proposed in [17]

**Cycle 1.** Cycle 1 initiated early industry engagement through a talk held at a virtual Danish meetup designed to raise interest among local practitioners. The core concepts of psychological safety and the results from [1] were presented to 70 attendants. Input was gathered through collaborative discussion activities held as part of the talk, as well as a following Q&A session. The meetup contained a call to sign up for Cycle 2’s workshop, which used the gathered input.

**Cycle 2.** Industry input for *tool* design continued through a digital workshop held on April 6<sup>th</sup> 2021 with 5 participants (14 sign-ups) comprising a mix of attendants from Cycle 1 and participants from industry. The goal of this workshop was to collect concrete experiences of psychological safety to inform *tool* design. It was conducted digitally using Zoom and Miro – an online collaborative white-board solution. Herein participants explored how psychological safety was experienced in their workplace, and proposed action points for making those experiences more psychologically safe. These points, and the discussions that emerged, became central to the design of *tools*. Concluding industry input collection, Cycle 2 resulted in the design of a *tool* compendium containing eight *tools* for working with psychological safety.

**Cycle 3.** The designed *tools* were evaluated to answer the research questions. Importantly, the subject of evaluation was the *tool* concept itself and the degree to which it aided the teams in working with psychological safety, not the success of each individual *tool* or their comparison.

Participating teams were recruited via calls to action distributed in several agile communities (e.g., AgilityLab: the host of the meetup from Cycle 1). These communities were primarily targeted for practical reasons: a large concentration of technical teams interested in processes and open to trying new ways of working. Four software teams from three different companies volunteered. Each team received a copy of the *tool* compendium, and chose their *tools* in an initial meeting with the researchers. Before implementation, teams were asked to conduct a shared viewing and open floor discussion of Edmondson’s Ted Talk on Psychological safety [7], in order to establish a baseline understanding of the subject.

Table 1: Tool Evaluation Survey Questions

Code	Question
Q1	How psychologically safe do you feel your team was prior to using the tools?
Q2	I felt psychologically safe while using the tools
TQ1	I enjoyed using the tool
TQ2	While using the tool, I reflected on things that my team does not normally discuss
TQ3	Using the tool made it easier for my team to work with psychological safety
TQ4	I could see the tool fit in with the way we normally work

*Note: All questions but Q1 were answered using a 5-point Likert scale; Q1 used a 7-point numerical scale for higher granularity. Questions TQ1-TQ4 were repeated for each tool.*

Teams then implemented their selected *tools* autonomously over a two-week period, immediately followed by two forms of evaluation: A) anonymous individual surveys distributed to all members of participating teams (see Table 1), and B) one-hour semi-structured group interviews held virtually with all members of each team. Both types of evaluation aimed to evaluate the degree to which the *tools* had worked as successful intervention activities, as well as their success of aiding the teams in working with psychological safety. The group interview focused on collecting this evaluation in the same group construct in which the psychological safety of the participating teams existed. The individual surveys allowed individual team members to voice their feedback through a safe medium wherein candid feedback could be given, even if their experience was negative or differed from that of the team. Following the conclusion of the two types of evaluation, results were gathered and analysed. Group interview responses were grouped using thematic clustering and analyzed alongside survey responses. The results of this analysis are presented in Section 5.

## 4 Building the Toolbox: Input and Design

This section will present the evolution of the artifacts designed in this paper, namely the *tools* for working with psychological safety. As mentioned in the introduction, this paper uses the term *tool* to refer to the tangible, descriptive representation of an intervention activity designed to affect change (intervene) on levels of psychological safety within a team. Concretely, a *tool* describes: A) an intervention activity, B) How and when this activity should be carried out, C) Meta-data about the activity, such as its duration or setting, D) Prerequisites of the activity, E) The purpose of the activity, and finally F) The expected outcome of the activity. Importantly, as the *tools* were designed during the Corona pandemic, all *tool* activities were designed to function within the boundaries of distributed and virtual work environments. The set of *tools* designed in this paper are collected and described in a "toolbox", namely the *tool* compendium, which is publicly available at [2]. In this compendium, each *tool* is presented alongside a short example of the *tool* in use. This format allows for a tangible representation of the interventions on psychological safety to exist in an accessible, shareable format, designed to enable any team (participants in this paper or otherwise) to

implement the *tools* autonomously without the researchers' involvement. This Section will not go into detail about the contents of each individual *tool*, but will rather aim to describe the four stages of artifact evolution through the four design science phases outlined in the Method Section 3, by presenting the four resulting artifact versions shown in Figure 1.

**AV0 – The Tool Concept.** *Tool* design was initialized by A) identifying the problem to be solved, namely the research questions put forward in Section 1, and B) defining the objectives of a solution to said problem; the designed artifacts (*tools* for working with psychological safety). Importantly, improving psychological safety was not a direct objective of the *tool* design process, which rather aimed to produce *tools* that enabled teams to work with psychological safety, potentially (hopefully) with the outcome of improving it. This distinction is important, as the improvement of psychological safety — a cultural change — is most likely to result from a team paying continuous attention to it over a longer period of time [9, Chapter 8]. It is therefore rather the aim that a *tool* successfully enables the team to work with psychological safety by creating a useful frame for this change process. Based on this objective, *tool* design began an iterative journey that continued throughout the following cycles. The initial inspiration began in the explorative work of the pre-study [1], wherein an early attempt at intervention on psychological safety was conducted. The learnings from implementing this intervention with industry software teams inspired both the problem to solve, and the artifacts designed to solve it. The goal of the design process was to synthesize research and industry experiences of psychological safety into an accessible but powerful set of intervention activities which teams could utilize to work with psychological safety, and to present these in a digestible format as *tools*. The word “tool” was chosen to present the activities as practical and tangible items as accessible as picking up a hammer to hammer in a nail. This was a core goal of *tool* design; using the *tools* should be as simple as possible, and should be compatible and useful regardless of a team's existing practices. This phase resulted in the first, early artifact version; the definition of a *tool* based on the objectives identified. As described earlier, *tools* were defined as tangible, descriptive representations of an intervention activity designed to affect change (intervene) on levels of psychological safety, allowing teams to pick up and implement them in their practice. The following three phases took this idea through iterative artifact design to realize this goal.

**AV1 – Tool Definition & Format.** To initiate *tool* design, the concept of psychological safety was broken down into several factors. Due to its complex nature, this would allow different *tools* to cover smaller subsets of the many aspects of psychological safety. This list of factors was synthesized by the researchers based on descriptions of psychological safety in Amy Edmondson's seminal work [6]. An additional factor of “awareness” (i.e, the awareness of the concept of psychological safety itself) was also added to this list, based on findings from the pre-study [1], in which an awareness workshop was conducted with positive results. The list of factors is presented in Table 2.

Table 2: Factors of Psychological Safety

Code	Factor	Description
F1	Awareness	The awareness of the concept itself.
F2	Identification	Identifying instances in your own particular environment.
F3	Asking Questions	Model curiosity, moving towards a culture of questions and interaction.
F4	Acknowledging Mistakes	Being able to detect, speak openly, and reflect on the mistakes of the team or the individual.
F5	Learning	Framing work as a learning problem, not an execution problem. Fostering a culture of experimentation.
F6	Challenging the Status Quo	Challenging the “way things are”, accepting constructive criticism, reflecting on existing process.
F7	Voicing Concerns or Ideas	Enabling team members to voice their concerns and ideas instead of holding back.

These factors would stay prominent throughout the further design evolution of the artifacts. They would come to influence the design of *tools* in phase 2 (see AV2 below), but for AV1 the factors were used to design the next step of artifact evolution: the *tool* one-page format, containing fields for different meta-data about the activity, such as when and why a team might use it, in addition to a description of the activity itself. This format was inspired by the “structure” concept of Liberating Structures, a collection of structures that provide “an alternative way to approach and design how people work together” [13]. The format was designed for use in an ideation workshop with industry participants, in which participants related the factors of psychological safety to their existing practice, and shared early ideas of intervention activities that were later used in *tool* design. The format used in this workshop additionally became the foundation for the presentation of *tools* in the final *tool* compendium.

**AV2 – Tool Design & Tool Compendium.** The second artifact version consisted of the design of the *tools* and their activities, based on the synthesis of collected input from industry and the research background of psychological safety. Industry input was gathered through the pre-study [1], a talk given at AgilityLab, and an ideation workshop with industry practitioners (see Section 3). Research input was drawn from literature on both psychological safety [6, 8], as well as agile practices and methods [4, 18]. Several *tools* were designed to be integrable with Scrum due to its popularity among agile practitioners. Eight *tools* were designed with the aim of covering the several aspects of psychological safety (see Table 2). Table 3 presents each of these *tools*, which factors of psychological safety they target, and where the inspiration for each *tool* was drawn from. For *tools* inspired directly by activities discussed in the *tool* workshop held with industry practitioners, the indicators WA (workshop activity) 1 through 5 are used. For *tools* wherein the inspiration was drawn directly from Edmondson’s descriptions [8] of how to work with that particular factor of psychological safety, the codes from the psychological safety factor table (Table 2) are used, pre-fixed with an E (i.e. EF1 for Edmondson’s descriptions of how to work with factor 1).

*Tools* were designed to differ along several axes of a design space in order to improve understanding of how teams could work with *tools* for psychological safety, as well as to provide a rich toolbox of viable options for the many different

Table 3: The Designed Tools - Factors and Inspiration

Tool	Factors	Inspiration Sources
What’s it to me?	F1, F2	Learning diary methods; Pre-study awareness workshop [1]
Checking In	F1, F3	Daily Scrum [18]; WA4; Edmondson’s 7-step survey for psychological safety [9]
Celebrating Mistakes	F2, F4, F5	EF4; WA5
Presenting the Journey	F2, F4, F5	EF4, EF5; Scrum sprint demo [18]; WA1
Three Questions	F3, F6, F7	EF3, EF7; Edmondson, impression management, risking appearing incompetent, intrusive and disruptive [8]
Acting on Concerns	F2, F6, F7	EF2, EF6, EF7; Scrum retrospective [18]
The Way Things Are	F2, F6, F7	EF6; Scrum retrospective [18]
Meeting From Hell	F2-F7	EF2-EF7; Liberating Structures (Triz) [13]; Scrum from Hell [19]

WA: Workshop Activity, F: factor of psychological safety (Table 2), EF: Edmondson’s Description of working with these factors

practices of different teams. Each *tool*’s placement within the design space axes was communicated in the *tool* compendium using an iconography, allowing teams to choose the *tools* they saw fit. Four axes were chosen for the design space:

**Setting.** The setting axis had two options: team activity or individual activity. Outside of practical differences, team activity *tools* could be more confronting, but allow for group reflection within teams finding such a setting useful, whereas individual activities could be a safer starting point other teams, or provide more time for individual reflection. Importantly, a team activity does not imply a physical meeting.

**Duration.** A linear scale of expected time needed to carry out a given *tool*’s intervention activity. Durations listed in the tool compendium were estimates made during *tool* design, and existed mostly to provide teams some expectation of time investment required. Letting *tools* vary across the duration axis allowed for the design of significantly different types of *tools*, ranging from short-and-sweet questions for a team to discuss, to longer activity formats.

**Frequency.** The frequency axis indicated the frequency with which a *tool* was expected to be carried out, and had the following values along its axis: once, iteratively (i.e., with a cadence of e.g. a week or a Scrum sprint), daily, and any. A value of “any” meant that the *tool* was used in an ad-hoc fashion, such as the *tool* “Celebrating Mistakes”, which involved addressing mistakes as they happened. Distributing *tools* along the frequency axis allowed for the design of *tools* that were either meant as incremental continuous improvement *tools*, to *tools* that were designed to be one-off conversation starters.

**Required Level of Comfort with Dissent.** The “Required Level of Comfort with Dissent” axis (numerical, 1-3) indicated how high a team’s comfort with dissent should be to achieve a constructive outcome from using the *tool*. While neither the scale nor a team’s self-assessment are well-defined values, distributing *tools* along this axis allowed *tool* design to challenge different teams at different levels, with self-assessment and *tool* selection being at the discretion of the teams. Some *tools* were designed to be introductory and safe, while others were more challenging. Importantly, comfort with dissent is a separate concept from psychological safety, though the two are related. A



Table 4: Overview of tools including selections from the evaluating teams

Tool	Setting	Duration	Frequency	Required Comfort	Selected by Teams
What’s it to me?	Individual	10min	Iteratively	1	1,2
Checking In	Group	15min	Iteratively	2	2,3
Celebrating Mistakes	Individual	10min	Any	2	None
Presenting the Journey	Group	15min	Iteratively	1	None
Three Questions	Group	5min + 5/person	Iteratively	1	None
Acting on Concerns	Group	20min	Iteratively	1	3,4
The Way Things Are	Group	10min + 3/person	Iteratively	1	1,2
Meeting from Hell	Group	45min	Once	3	1,3,4

team could struggle with some factors of psychological safety, such as voicing concerns or challenging the status quo, but still have a strong comfort with dissent whenever dissent occurs. Such a team might have mediocre psychological safety, but might still be in a position to get a constructive outcome from *tools* with a higher requirement of comfort with dissent.

An aim of this design process was to spread *tools* across the design space, providing both safe and challenging options that could fit different practices. The only area of the design space for which no *tools* were designed, was the combination of short duration and a high requirement for level of comfort with dissent. This design decision was made to avoid exposing teams to challenging activities without being given the proper time to engage and reflect. For the purposes of sharing the designed *tools* for implementation, they were collected in a single document; the *tool* compendium. This compendium contained all of the designed *tools*, as well as introductions to the concept of psychological safety and using the *tools*. The compendium was designed with the aim that any team could pick up the compendium and use the *tools* autonomously, without any interaction with the researchers. This version of the designed artifact — the *tool* compendium — was the final artifact version used in evaluation.

**AV3 – Finalised Tool Compendium.** During the evaluation of AV2, several points were brought up resulting in minor changes being made for future users of the *tool* compendium. Upon the conclusion of evaluation, it was also decided that the introductory activity of watching Edmondson’s Talk on Psychological Safety [7] would be added as the ninth *tool*, giving future *tool* compendium users a similar introduction to the subject, as the one given to the participating teams in this paper. This is also supported by Google’s similar recommendation of the talk in [5]. This final version (AV3) of the *tool* compendium can be found in [2].

## 5 Results

This section presents the results from *tool* evaluation. *Tools* were evaluated with four software teams of 9, 6, 4, and 3 members from three different SaaS companies working with variations of Scrum. Table 4 presents the characteristics of the designed *tools* and details which team selected them for implementation.

Table 5: Survey Answers

Answer	TQ1	TQ1*	TQ2	TQ2*	TQ3	TQ3*	TQ4	TQ4*
Positive	58%	67%	71%	81%	64%	69%	56%	66%
Neutral	27%	31%	18%	15%	24%	25%	20%	23%
Negative	15%	2%	11%	4%	13%	6%	24%	11%

Note 1:  $TQx^*$  columns are presenting results disregarding the answers from Team 3 and 4 in relation to Meeting from Hell.

Note 2:  $N(TQx) = 55$ ;  $N(TQx^*) = 48$ .

## 5.1 Survey Results

Table 5 presents survey responses, grouped as positive (agree + strongly agree), neutral, and negative (disagree + strongly disagree) responses, for ease of presentation. All teams reported a high level of psychological safety prior to using the *tools* (Q1 between 5.8 to 6.75, 7-point scale). Overall, teams expressed enjoyment (TQ1), positive reflection (TQ2), and engagement with psychological safety (TQ3) across all tested *tools* and were mostly positive regarding the likelihood of fitting the *tools* in their process (TQ4). A notable pattern in the results was the exposure to the *Meeting from Hell* tool. While for Team 1 the use of this tool was still generally positive, for Team 3 and 4, the use of *Meeting from Hell* was a negative experience and the majority of the negative responses received in the survey are related to these pairing of team and tool. Table 5 accounts for this pattern by presenting two versions of response data: TQx for the overall and TQx\* disregarding answers of Team 3 and 4 in relation to Meeting from Hell.

## 5.2 Evaluation Interview Results

The evaluation group interviews were held with each participating team. Each session was annotated and recorded. Thematic clustering was used to analyse annotations and recordings, which led to the six themes presented below.

**Aiding teams in working towards better psychological safety.** Teams were extremely positive on this topic. Participants stated that the *tools* (with the exception of some experiences with the tool *Meeting from Hell* discussed later) they used enabled constructive discussions about psychological safety, which they might not have had otherwise. “*I think that it was good for the team. It made us discuss stuff that we don’t usually discuss.*” says team 2. While team 4 highlights how “*Acting on Concerns made us have a lot of good discussions [...] I feel like we talked about it in a new way. Hopefully it would have come up anyways, but it was good to get it out in the beginning of the project.*”

Multiple teams also experienced process improvements during their participation. While this was not a direct goal of this paper, the ultimate purpose of improving psychological safety is that of team excellence, not just pleasant culture [8]. Team 1 reports that “*it has provided some efficiency to our meetings, and some afterthought to one self.*” Team 2 continues: “*The result of our The Way Things Are was really good. It actually already feels like it’s made a bit of*

*a change in how we do our stand-ups. [...] I am actually confident now, that no one is sitting and struggling with something, because we actually mention it.”*

Finally, participants indicated that the *tools* were engaging and functional team activities. According to team 2, “[...] it’s quite often that our discussion go more to one domain than the other [...] But actually, for all of our tries with the tools, I noticed that everyone participated, all the way through.”

**Putting a label on it.** Several participants spoke to the concept of psychological safety being a label to several things they had either worked with or otherwise experienced in the past and that having a name for this concept was almost as helpful as the *tools* themselves. This finding is in line with the experiences of the awareness workshop conducted in [1], in which some participants experienced higher levels of psychological safety after awareness of the concept was spread within the participating teams. Team 2 confirms that they were “*really good conversation starters in the sense that it’s not necessarily things that are easy to bring up normally, but putting it within a frame made it very easy to go about.*” And also: “*have it named within a team, right. We talked about this, we talked that it’s okay to bring it up*”. Interestingly, for team 3 “*it is clear that the idea of speaking about psychological safety is something we might want to do*”, and team 1 explains how while “*we are free to challenge things already, [...] I still think that [using tools] can be a good jump start for some people.*”

**Prompted with a purpose.** Another re-occurring theme among participants was that of simply taking the *tools* as a prompt to have a discussion, which they might already have been able to have, but were not having. One reason as to why the teams did not have these discussions, was described as trying to avoid appearing a certain way to your co-workers, something that Edmondson identifies as key reason why people hold back, namely because of impression management [9]. When prompted to purposefully engage in this kind of behaviour, participants expressed that this worry was easier to let go, especially when seeing other team members engage in similar behaviour. “*Sometimes*” – says team 4 – “*if you are speaking about concerns, you might seem like a grumpy old man that is only seeing issues and road blocks, but actually [pause] making this room where you map out all the different concerns, and see that other people have the same concerns, or talk about some of the things that you believe are concerns which is not a concern for others. I think it’s just a great tool.*”

Others simply had not found a space for these discussions, or did not know where to start. Team 4 says that *[the tool] is just great at facilitating and getting those questions asked.*, which is confirmed by Team 3 that states: “*Acting on Concerns is a great way to kind of create a space, where [psychological safety, concerns] is what you are speaking about. And that just provides insane amounts of value. That is at least how I experienced it with everyone.*” Additionally, acknowledging that working with psychological safety was worth allocating time for was identified as another enabling factor. Team 1 reports how “*it was great to see that we take it seriously, that we look into psychological safety, that we put it on our agenda, and that we want to spend time on it.*”

**Does it matter what tools we use?** During interviews, several participants pondered whether the overall outcome of implementation could differ depending on the *tools* selected. While the concrete experiences with each *tool* differed, and some *tools* were preferred over others, several teams, like Team 1, expressed that “*it almost does not matter what tool you use*”, alluding to the strength of simply addressing the topic of psychological safety. This could indicate that, when the *tool* activity goes well, a successful *tool* leaves the focus to the team’s self-reflection rather than the *tool* itself. However, as mentioned earlier, some teams (i.e., Team 3 and 4) did have negative experiences with one of their *tools*, *Meeting from Hell*. Team 3 described the *tool* as “*decidedly awkward*”, struggling with getting the discussion started as “*it requires a lot from the person hosting it*”, who needs to “*assume control for it to go well*”. Team 4 also reported that their negative experience might have been due to a “*wrong mix of personas*”. Given that Team 1 had a very different (positive) experience with *Meeting from Hell*, a poor fit between a team and the *tool* could explain a negative experience. Additionally, Team 3 and 4 being from the same company might have been related to their similar experience. Team 3 and 4 successfully implemented their other *tools* explicitly voicing their preference: “*I don’t think that Meeting from Hell is a particularly bad exercise [...] but it didn’t create a lot of value considering the time we spent on it, whereas Acting on Concerns created a lot of value and a great discussion and dialogue with less effort*” (Team 4).

**The impact of existing levels of psychological safety.** The question of how a team’s existing level of psychological safety might impact *tool* outcomes was discussed by several teams. Participants reflected on whether a team with a lower existing level of psychological safety would have benefited more than one with a very high level, and whether a team with a “high enough” level of psychological safety would benefit from using the *tools* in the first place. These discussions resulted in similar assessments across teams: “*discussion about [psychological safety] is never bad, even if [the level of psychological safety] might still be good beforehand*” (Team 2); Team 1 “*did not think that [psychological safety] was a big issue [but] it was great to see that we take it seriously, that we look into psychological safety, that we put it on our agenda, and that we want to spend time on it*”; and, Team 2 highlights that an individual might think “*oh yeah, this place is super psychologically safe’, when in reality my team members are just shitting themselves if they have to say anything.*”

**Future use of Tools** As the final step of the evaluation interviews, teams were asked if they could see themselves using their *tools* again in the future. All teams responded positively with at least one *tool* they would like to continue to use, while some teams identified wanting to use multiple. Team 1 describes how “*The Way Things Are was super. It is a good tool. [...] we could definitely [use it again]. And also Meeting from Hell. [...] I think I could see Meeting from Hell in a [company name] version, wherein you take it up once in a while.*” Team 2 thinks that “*we should do another The Way Things Are. Not necessarily the next, like, week or month or anything, but eventually. I think that was a really fun experience. [...] I definitely think it could be interesting to try it again.*” And,

even more decisively regarding *Acting on Concerns*: Team 3 “*I am convinced that we will be using it again*”; and Team 4, “*I think it is just a great tool. It is definitely something we will use again, I believe, in all our big projects, actually.*”

## 6 Discussion

This section will discuss the results presented in Section 5. Results are discussed per research question in the subsections below, followed by future work. Where survey results are referenced, two results are presented using the following format: 25% (35%), parallel to the format of the results presented in Table 5, showing results from TQ<sub>x</sub>, and TQ<sub>x</sub>\* respectively.

### 6.1 RQ1: Designing Tools to Enable Agile Software Teams to Work with Psychological Safety as Part of Their Practice

This paper saw *tools* for working with psychological safety designed as the synthesis of research and industry input through an iterative process using design science (see Section 3). These *tools* were implemented and evaluated with four industry software teams. In evaluation surveys, 64% (69%) agreed that using the *tools* made it easier for their teams to work with psychological safety, and 58% (67%) enjoyed using the *tools*. For a potentially sensitive subject such as psychological safety, the teams enjoying using the *tools* is an important aspect of whether those *tools* can aid the teams in working towards better psychological safety, especially for continuous use. Evaluation interviews saw overwhelmingly positive responses, with participants identifying the *tools* as enabling them to have discussions they did not normally have, and finding it easier to speak up. Additionally, 56% (66%) reported that they could see the *tools* they used fit their existing practice. These results indicate that the designed *tools* were largely successful, answering the research question of how such *tools* can be designed; namely through the synthesis of research on psychological safety, and the experiences of industry practitioners, into bite-sized intervention activities, shared through one-page descriptions, using the *tool* format (see Section 4).

The *tool* concept itself seemed to provide a useful frame for working with psychological safety for the teams. The presented format and the design space created for the *tools* appeared to make the different *tools* understandable and easy to pick up and implement for the teams, with none of the teams having any facilitation being conducted by the researchers. This indicates that the *tool* concept was successful, and could be re-used for the design of future *tools*.

### 6.2 RQ2: Aiding Agile Software Teams in Working Towards Better Psychological Safety through Tools

The primary aim of the designed *tools* was to aid software teams in working towards better psychological safety. While the *tools* themselves could not guarantee the improvement of psychological safety within the teams directly, *tools*

were designed to make it easier for teams to achieve this goal by providing an enabling frame for the team to work within. Most participants 72% (81%) reported that using the *tools* caused them to reflect on things which their team did not normally discuss. From interviews, participants reported that they in some instances found it easier to speak up and voice their concerns during or after using the *tools*, and recounted experiences in which they had spoken up as a direct result of using a *tool*. Even teams that viewed themselves as having high psychological safety prior to using the *tools* reported that they felt more confident in their psychological safety after using the *tools* within their team. Several participants mentioned that thinking that your team has a high level of psychological safety is different to openly discussing and aligning individual perceptions with the team. Additionally, participants identified that the *tools* gave their teams a needed prompt to address unspoken subjects. Allocating the time to discuss these things as a team was deemed an important part of the successful experience, with some participants stating that they found the prompt and the time allocation even more impactful than the activities of the *tools* themselves.

All participating teams reported that they wanted to continue using one or more of their selected *tools* going forward, in order to continue working with psychological safety. This both indicates a positive experience using the designed *tools*, as well as an expressed interest in continuous attention being paid to psychological safety over time, using these *tools*. This outcome falls in line with Edmondson's descriptions of psychological safety requiring continuous renewal over time [9, Chapter 8], further indicating that the designed *tools* could continuously aid software teams on their journey of working with psychological safety.

### 6.3 Threats to Validity

**Team Levels of Psychological Safety** In evaluation surveys, all teams unanimously reported high existing levels of psychological safety. This bias could be explained by the nature of participation, namely teams from an Agile community that actively decided to opt in. However, it raises the question of whether the success of the designed *tools* depends on the existing levels of psychological safety of the implementing teams. This is a complex question, as objective measurements of psychological safety have had limited success [1, 16], rendering existing levels of psychological safety an undefined metric for most teams. To accommodate for this uncertainty, the designed *tools* were distributed across a varied design space, allowing different options for different teams. However, the question of how teams with little to no psychological safety can initiate their journey with psychological safety was considered out of scope of this paper, as it is likely to require a specific focus on such environments.

**The Tool or The Toolbox?** The center point of both design and evaluation was the *tool* concept itself, its design space, and the degree to which *tools* implementing it could be integrated into the practice of software teams. As mentioned in Section 3, the focus was on the *tool* concept, and not necessarily on the success or differences of the intervention activities of individual *tools*. This choice had several implications, such as *tool* selection being conducted with team and

*tool* fit being prioritized, rather than aiming for all *tools* to be selected. Additionally, the autonomous selection and implementation of different *tools* among the individual teams most likely resulted in differing experiences of individual *tools*. This is, however, a direct goal of *tool* design; namely that of finding a way for software teams to work with psychological safety, regardless of *tool* selection, practice or implementation details. The *tools* were by design not prescriptive, aiming rather to provide guidelines for teams to engage with the concept of psychological safety, than exact rules of implementation or discussion. To this end, the evaluation shows that the designed *tool* concept is one such way for teams to work with psychological safety as part of their practice. Whether more successful *tools* can be designed within the design space, or indeed the design space itself can be improved, is a topic for future research.

**Tool Implementation** The designed *tools* were implemented over a two-week period by participating teams. While it is possible that a longer duration could provide richer data, the intent of this paper was to experiment with integrating working with psychological safety into the practice of agile software teams. For this reason, many of the *tools* were designed around common foundations of agile practices, such as iterative structures, and had their frequency of use in part defined by such iterations. As such, it was the aim to explore the insertion of the designed *tools* into an existing iterative structure, which for the participating teams aligned with the two-week implementation period. Given the results of this paper, continuous implementation and evaluation could provide further insights.

#### 6.4 Future Work

Research on psychological safety is still very new to the domain of software. The work conducted in this paper is an initial step into a broader subject of how software teams can adopt, work with, and improve their psychological safety. The continuous implementation and evaluation of the *tool* concept is a natural continuation of this paper. For continuous evaluation of the effect of *tool* usage on psychological safety over time, repeated quantitative measurement techniques akin to those designed by O'Donovan et al. [16] (as was utilized in the pre-study [1]) could be useful.

## 7 Conclusion

Using design science research, this paper presents the design of actionable *tools* to aid and enable software teams in working with psychological safety. Eight such *tools* were designed and implemented autonomously by 4 software teams over a two-week period, followed by survey and group interview evaluations. Evaluation showed that teams found the *tools* both enjoyable and helpful as both conversation starters and frames within which to work with psychological safety. Teams additionally found the *tools* to fit within their existing practice, and universally planned to use one or more of their *tools* in the future.

## References

1. Agerlin Christensen, M.: Tr-2021-207: Measuring and affecting psychological safety in software teams: A case study. Tech. rep., IT University of Copenhagen (2020)
2. Agerlin Christensen, M.: Tool compendium: Tools for working with psychological safety in agile teams [online] (2022)
3. Aken, J.E.v.: Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules. *Journal of management studies* 41(2), 219–246 (2004)
4. Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., et al.: Manifesto for agile software development (2001)
5. Dubey, A.: Project aristotle - understanding team effectiveness (2014), <https://rework.withgoogle.com/print/guides/5721312655835136/>
6. Edmondson, A.: Psychological safety and learning behavior in work teams. *Administrative science quarterly* 44(2), 350–383 (1999)
7. Edmondson, A.: Building a psychologically safe workplace (2014), <https://www.youtube.com/watch?v=LhoLuui9gX8>
8. Edmondson, A.C.: Managing the risk of learning: Psychological safety in work teams. Citeseer (2002)
9. Edmondson, A.C.: The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth. John Wiley & Sons (2018)
10. Forsgren, N., Smith, D., Humble, J., Frazelle, J.: 2019 accelerate state of devops report (2019)
11. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. *MIS quarterly* pp. 75–105 (2004)
12. Lenberg, P., Feldt, R.: Psychological safety and norm clarity in software engineering teams. In: Proceedings of the 11th international workshop on cooperative and human aspects of software engineering. pp. 79–86 (2018)
13. Lipmanowicz, H., McCandless, K.: The surprising power of liberating structures: Simple rules to unleash a culture of innovation. Liberating Structures Press Seattle, WA (2013)
14. O'Donovan, R., McAuliffe, E.: Exploring psychological safety in healthcare teams to inform the development of interventions: combining observational, survey and interview data. *BMC Health Services Research* 20(1), 1–16 (2020)
15. O'Donovan, R., McAuliffe, E.: A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour. *BMC health services research* 20(1), 1–11 (2020)
16. O'Donovan, R., Van Dun, D., McAuliffe, E.: Measuring psychological safety in healthcare teams: developing an observational measure to complement survey methods. *BMC medical research methodology* 20(1), 1–17 (2020)
17. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. *Journal of management information systems* 24(3), 45–77 (2007)
18. Schwaber, K., Sutherland, J.: The scrum guide. Scrum Alliance (2020)
19. Wake, B.: Scrum from hell - xp123 (2004), <https://xp123.com/articles/scrum-from-hell/>
20. Wieringa, R.: Design science as nested problem solving. In: Proceedings of the 4th international conference on design science research in information systems and technology. pp. 1–12 (2009)



## Acknowledgments

- Write intro focus on design science framing
- Background ... design science
- Participant demographics, company sizes/types (startup? Small team in bigger org?)
- Participant demographics. what about the process they follow. some words on this
- Decide in and out for results (based on DSRM-framing of RQs)
- Engstrom 2020 will be used for terminology (and framing/description of tools) but
- the techno rule will not be applied to a problem instance but rather to an objective-centered approach in the DSRM model
  
- agile or scrum? we chose the participants from an agile setting .. how do you use tools
- ... we worked as a starting point in an agile community
- ... we keep agile, this needs an explanation
- is DK a special case? needs a discussion ... in scandinavia human resources are taken care of .. the barrier to entry is culturally lower
- can these tools be applied to any tool? see thesis . generalizability
- how do you know you have too little PS? maybe out of scope ... people faking PS
- ISBN 978-87-7949-373-5 ... TR-2021-207
- charts onz lucid.app
  
- Technical Report publication of the work with Allan. Change the data visualization of survey answers (use the visualization from the thesis)
- Inclusion of the outcome of tool testing? (screenshots)? If not for paper, use in presentation
- E-mail participating companies about their methods used
- *inkscape --export -type = pdfsvg \* svg*
- link to dsrm [https://lucid.app/lucidchart/de45f0e3-e318-4a09-aa69-23248abe0b7b/edit?invitationId=inv\\_5d3b5ea7-7d4c-425a-910b-b76a9c381438&page=0\\_0#](https://lucid.app/lucidchart/de45f0e3-e318-4a09-aa69-23248abe0b7b/edit?invitationId=inv_5d3b5ea7-7d4c-425a-910b-b76a9c381438&page=0_0#)
- In addition to the survey question answers, one participant entered the below comment in the open feedback field at the end of the survey.  
Presenting psychological safety bundled in nice, bite-sized packages made it tangible and easy to work with, so good job!  
(Team Member, Team 2 - May 2021)
  
- full (15 pages)
- Abstracts due: Feb 7, 2022
- Submission deadline: Feb 14, 2022
- Notification to authors: Mar 25, 2022
- Camera-ready due: April 15, 2022
- All dates are defined as AoE (Anywhere on Earth).

TODO LAST MINUTE:

- check if reference on Figure 1 to point to [1]

Feb 4 TODOS:

- Discuss figure 1
- Check references
- Figure 1 red text

TOD: Two quotes we never used from "any other feedback or comments?"

Presenting psychological safety bundled in nice, bite-sized packages made it tangible and easy to work with, so good job!

mainly that the naming of the tools is a little confusing. Otherwise very positive about this experiment!

XP Review TODOS:

- Fig. 1 problems ("the sequence of the activities depicted in Fig. 1 didn't match the description text. For example, the talk at AgilityLab happened at the beginning of Cycle 1, not at the end of it? Similarly in Cycle 2, the workshop with industry practitioners provided input to the development of tool compendium, not evaluating it? Please check Fig. 1 and the text to make them consistent")
  - Threats to Validity (NEW SECTION) - look for a DS paper that does TtV
    - Tools not selected (why did selection vary?)
    - High existing self-reported level of PS (Are the tools only for mature teams?)
    - Team implementation of tools (especially with no observation), could be very different.
    - Survey respondents remembering what PS is ...Eh
    - Duration of implementation (2-week)
  - Team Demographics
    - Details on team size?
    - Team roles?
    - Team Demographics in method?
- I am not sure about how much depth to go into here.
- What does "working" with PS mean? (Improves discussions? Use more suitable word than "working", or include practices) Minor edits to introduction to include practices. Maybe there's a better word?
  - Evaluation
    - Interviews: online/in person? participants? Clarified in method section
    - When were surveys distributed? Clarified in method section
    - Participants understanding PS (definition) while answering surveys ...Eh
    - Thematic analysis and Likert scale choices given in Results, not method
- Short note mentioning thematic clustering at the end of method
- Two versions of answers because "one team did not like a tool"
  - Pre-study is not a formal publication, credibility?
  - "Factor" of psychological safety the right word?