

*This is a post-print version. This article may not exactly replicate the final version published in the journal. Once the typeset version is finished, the final peer-reviewed and edited copy of this manuscript can be found at the homepage of the Small Group Research:*

<https://journals.sagepub.com/doi/full/10.1177/1046496420904126>

**Citation:**

Klonek, F.E., Meinecke, A., Hay, G., & Parker, S. (in press). Capturing team dynamics in the wild: The communication analysis tool. *Small Group Research*.

**Capturing Team Dynamics in the Wild: The Communication Analysis Tool**

Florian E. Klonek<sup>1</sup>, Annika L. Meinecke<sup>2</sup>, Georgia Hay<sup>3,1</sup>, and Sharon K. Parker<sup>1</sup>

<sup>1</sup> Future of Work Institute / Center for Transformative Work Design, Business and Law School/ Curtin University, Perth, Australia

<sup>2</sup> Industrial/Organizational Psychology, Faculty of Psychology and Human Movement Science, University of Hamburg, Hamburg, Germany

<sup>3</sup> Business School, University of Western Australia, Perth, Australia

**Author Note**

This research was supported by funding from the ARC Australian Laureate Fellowships “Transformative work design for health, skills and agility” (FL160100033), and funding from the German Academic Exchange Service and Australian Universities (project-ID 57388350).

Correspondence concerning this article should be addressed to Florian E. Klonek, Centre for Transformative Work Design, Curtin University, Western Australia.

E-mail: [florian.klonek@curtin.edu.au](mailto:florian.klonek@curtin.edu.au)

**Acknowledgements**

We would like to thank Gareth Baynam (the Program Director of the Undiagnosed Diseases Program, UDP-WA), Lauren Dreyer (Genetic Services), Stephanie Broley (Genetic Services), Alicia Bauskis and Hugh Dawkins (Department of Health) for their invaluable support of our research. We also want to express our gratitude to all participants involved in the UDP-WA program for allowing us to carry out observations in this very sensitive context. We would like to sincerely thank Ding Wang who has integrated all technical features into the CAT software and has been driving the development over the past 2.5 years. Finally, we want to acknowledge the work from our research assistants Natasha Zint, Dale Long, and Meredith Carr who provided invaluable support in coding the research data.

### Abstract

Capturing team processes, which are highly dynamic and quickly unfold over time, requires methods that go beyond standard self-report measures. However, quantitative observational methods are challenging when teams are observed *in the wild*, that is, in their full-situated context. Technologically advanced tools that enable high-resolution measurements in the wild are rare and, when they exist, expensive. The present research advances high-resolution measurement of team processes by introducing a technological application—the Communication Analysis Tool (CAT)—that captures fine-grained interactions in real workplace contexts. We introduce four core features of CAT: (1) customized coding measures, (2) session-based feedback on interrater reliability, (3) visualization and feedback options for displaying team dynamics, and (4) an export function to conduct advanced statistical analyses on effective team processes. We illustrate these core features using data from an organizational field project on multi-disciplinary teams tasked with diagnosing patients with uncommon and highly complex medical conditions.

*Keywords:* behavior coding, interaction analysis, open-source software, interrater reliability, temporal dynamics, behavior observation software

### **Capturing Team Dynamics in the Wild: The Communication Analysis Tool**

In-depth analyses of team processes can help scholars to understand the temporal dynamics of team behavior that arise in interdependent situations. Team processes are dynamic, unfold over time, and can sometimes change quickly (e.g., Leenders et al., 2016; Kozlowski, 2015; Schechter et al., 2017). As an example, take the Miracle on the Hudson; a crisis situation that forced the crew of the US Airways Flight 1549 to land on the Hudson River (New York, USA) after the plane's engines failed due to a collision with a flock of geese (NTSB, 2010). Within a period of less than 5 minutes, and under conditions of high stress, the team had to quickly decide and communicate about what to do (e.g., by identifying and declaring the incident, assigning roles and responsibilities, and maintaining the chain of commands). Understanding how these time-dependent interactions unfold and change moment to moment—allowing the team to successfully adapt to the crisis at hand—can provide crucial learnings and contribute important knowledge for similar team emergencies (Eisen & Savel, 2009).

Over the last decade, we have seen a growing interest in understanding the micro-dynamics underlying team functioning (Cronin et al., 2011; Kloneket al., 2019; Kozlowski et al., 2013; Schechter et al., 2017). Efforts are being made to explore team process dynamics not just in controlled laboratory settings (e.g., Kennedy, & McComb, 2014; Uitdewilligen et al., 2016) but also within their actual organizational and sociotechnical context, which has also been labeled as studying teams “in the wild” (see Salas et al., 2008, p. 544). Examples of such contexts include, but are not limited to, flight crews, power plant operating teams, medical teams, and crisis management teams (e.g., Lei et al., 2016; Stachowski et al., 2009; Schmutz et al., 2015; Uitdewilligen & Waller, 2018).

To adequately map the process dynamics in these teams, scholars have recommended the use of movie-like temporal resolution approaches to allow for near continuous assessments of social processes (Kozlowski, 2015; Leenders et al., 2016). Although video-based and related observational measurement approaches (e.g., based on sensor technology) are capable of capturing repeated measures of behaviors, and hence enable the assessment of micro processes, studying team dynamics is challenging (Lehmann-Willenbrock & Allen, 2018). First, fine-grained observational research is labor intensive because observations need to be systematically documented to obtain a structured dataset. Professional software solutions can be of help to the researcher as they enable the consistent coding of behaviors over time. However, most commercial software products that support such data collection efforts come at high costs and do not necessarily meet the researcher's requirements (Klonek et al., 2015; Lehmann-Willenbrock & Allen, 2018). For example, commercial software like Noldus Observer XT (Noldus et al., 2000) or Mangold Interact that are frequently turned to for coding and analyzing video recordings have expensive licenses. Second, existing technological solutions are limited in terms of providing immediate and intuitive visual feedback to participants (or to the researcher), or providing feedback on data quality (e.g., interrater reliability) that limits the possibility for quick insights and actionable interpretations. Third, commercial software solutions are not well suited to support collaborations across different locations or laboratories. Researchers typically have to install local computer programs and work with software-specific data files that cannot be opened in conventional programs. This can complicate collaborations by restricting opportunities for geographically dispersed work. For example, in a field project with a multi-national organization, on-site observers should be able to collect and log observational data while cooperation partners have direct access to digital data saved in a repository.

We are not aware of any existing technological solution that currently fully addresses these challenges. While open-source solutions exist, they often have idiosyncratic limitations. For example, the *Observational Data Coding System* (Maclin & Maclin, 2005) and *The Simple Video Coder* (Barto et al., 2017) are restricted to a limited number of coding categories (e.g., they only allow between 10 to 18 categories), cannot be used for live observations, and require knowledge of complex programming language.

The goal of this methods spotlight is to introduce a browser-based software solution that captures process dynamics in groups and teams, either via video/audio recordings or via live observations. Specifically, we designed a tool—the Communication Analysis Tool (CAT)—that allows researchers to integrate custom-made coding and rating measures in a user-friendly way (i.e., requiring little knowledge in software installation and/or statistical analyses), allows researchers to assess data quality with respect to reliability, and allows researchers and participants access to immediate visual feedback about coded team processes. Overall, CAT has the potential to help collect fine-grained moment-to-moment team dynamics, spark novel research questions about team processes in real-life teams, and aid in behavioral interventions based on immediate feedback about team process dynamics (defined as the pattern of changes in a phenomena over time; Roe et al., 2008). The tool presented here is newly developed and has not yet been intensively tested. The aim of this article, therefore, is to present the manifold possibilities of the tool and its functionality, in order to pave the way for future (cross-disciplinary) research.

The remainder of this article is structured as follows. We first describe the organizational field research context that stimulated the development of the CAT software. Following this, we provide a short overview on quantitative group interaction analysis which is the type of

methodology that aligns with a tool like CAT. Doing so, we also review existing software options in this space. We then outline the four core features of the CAT software in more detail; (1) customized coding measures, (2) interrater reliability, (3) feedback and visualization of team dynamics, and (4) an export function to triangulate the data with other measure and/or to carry out advanced statistical analyses for testing research hypotheses. We illustrate each core feature with an example from our own research project (multi-disciplinary problem solving in health care teams). Finally, we reflect on current limitations of the software and discuss practical implications of the tool.

### **Research Context: Development of the CAT software**

We developed CAT as a field research tool in collaboration with a program aimed at improving medical diagnosis for uncommon diseases in children. The program introduced multi-disciplinary expert panel meetings (Baynam et al., 2017; Koole et al., 2017; Oborn & Dawson, 2010) to improve patient care and find answers for patients with highly complex, rare, and long-standing medical conditions. Our project was concerned with improving and understanding the complex problem-solving processes that occur during these monthly expert panel meetings. Patients are referred to the program through the Genetic Services or a health clinic when they meet all of the following participation criteria: the child is at least six months old, the child has chronic and complex health problems that affect multiple body systems, the child has no diagnosis, and the child has a history of multiple hospital admissions and specialist assessments. The overarching goal is to identify a diagnosis that can explain the complexity of a patient's symptoms. The complex nature of these rare diseases creates a disproportionately large impact on the public health system and a substantial financial and psychological impact on the patients (Walker et al., 2016).



After referral to the program, patient-related information is reviewed in the monthly expert panel meetings. These interdisciplinary meetings are attended by specialists from multiple disciplines of the medical field, including clinical genetics, neurology, imaging, endocrinology, gastroenterology, cardiology, hematology, ophthalmology, respiratory medicine, metabolic medicine, and others. All members receive a summary of the patient's medical history before the meeting; during the meeting members discuss their ideas and engage in interactive problem. A typical meeting lasts for about 1 hour and is attended by 6 to 12 members. The core task of these meetings is to identify a potential diagnosis, discuss possible medical pathways, and make decisions on further clinical assessments. The meetings typically conclude with a decision on which (further) genetic tests are to be performed, and whether and what further non-genetic investigations and assessments are required. Finally, the panel decides whether or not to share the data with an international collaboration system. Since the introduction of the expert panel meetings, the diagnostic rate of patients going through the program has significantly increased (nearly doubled, to 55%) when compared to the previous approach (Baynam et al., 2016).

After obtaining ethical approval, we were given access to join the expert panel meetings. In this context, we developed the CAT software to gain a better understanding of the team process dynamics that occur within these panel discussions (the website contains more details on the development of CAT). The leadership team expressed a general interest in understanding how the meetings could be further improved in order to increase the diagnostic success rate. In the early stages of this project, our data collection relied on live observations of the meetings. Because sensitive patient data was shared and discussed during the meetings, the program's leadership wanted to create a psychologically safe environment in which meeting participants feel free to voice unusual ideas. It was believed that recording meetings from the very beginning

of the program could be counterproductive to generate the desired atmosphere and to establish trust. After sufficient trust was built, we also had the opportunity to record meetings on video. Hence, CAT was first and foremost developed as a tool for live observation. Functionalities to analyze media files (i.e., video/audio-files) were added at a later stage. We developed CAT with the goal in mind that it would allow observers to capture crucial moments (i.e., in terms of identifying crucial time points) of sharing critical knowledge, assess and quantify important team processes (e.g., asking questions, team learning) for each meeting, and to feedback this information to the leadership team.

### **A Primer on Quantitative Group Interaction Analysis**

Before we start outlining the specific features of the CAT software, we provide readers with an overview of group interaction analysis<sup>1</sup> to help readers see how our software tool fits into this growing research field. Interaction analysis involves a set of systematic techniques or steps to make valid interpretations from observations of naturally occurring interactions (Keyton, 2018). Central to interaction analysis are coding schemes that give structure to the observations.

#### **Coding Schemes: Enabling Systematic Observation**

Researchers with a quantitative focus use *coding schemes* to quantify specific team behaviors (representing a particular team construct; for overviews see Keyton, 2018; Waller & Kaplan, 2018). We designed CAT for researchers with such quantitative focus in mind. When collecting group or team process data, quantitative researchers use external observers as *coders*. This can be the researcher who developed the particular research question at hand and/or trained research assistants. *Behavior coding* (or annotation) means that these trained observers assign behavioral codes (e.g., asking a question) to discrete units (e.g., speaker turns or events that express a complete thought) by using a predefined coding scheme, such as the well-known

Interaction Process Analysis (IPA) coding system by Bales (1950). Similarly, *behavior rating* entails the use of a predefined rating scale to assess the extent/ or quality of a group phenomenon within a specific time window of observation (e.g., “To what extent did the team engage in idea exploration during the last 5 minutes?” 1 = *not at all*, 5 = *a great deal*). Rating typically involves the use of Likert-type scales and focuses on larger time frames in comparison to behavior coding. As such, behavior rating makes use of observers’ aggregated judgements and is especially useful for (team) constructs that are more socially bound (e.g., Meinecke et al., 2016). Both approaches can be implemented within CAT. In the following, we focus on the application of behavioral coding (and not rating), as this was central to our research focus.

Despite a multitude of available coding schemes (see Brauner et al., 2018, for an overview of existing group interaction coding schemes), researchers often have to adapt existing measures to align them with the specific requirements of their study context. For example, a coding scheme developed for creative problem solving in student laboratory teams might not be well suited to capture creative behaviors occurring within organizational teams (see also Luciano et al., 2018). As a result, coding schemes often evolve over time and a critical function of any computer-assisted coding tool is its ability to allow further refinements or adaptations. We kept this need for flexibility in mind when we developed CAT and tried to find the right balance between offering a coding tool that guides researchers through the necessary steps in conducting team process research and yet, at the same time, allows making adjustments as needed.

### **Unitizing: Parsing the Stream of Behavior**

Researchers using group interaction analysis and who carry out behavioral coding need to decide on a sampling plan. A common distinction is made between *timed-event* (or simply event) sampling versus *interval* (or time) sampling (Bakeman & Quera, 2011). In a timed-event

sampling plan, coders assign codes based on parsing rules (frequently also referred to as unitizing rules) which are specified by the respective coding measure (e.g., Schermuly & Scholl, 2012; Keyton, 2009). The general idea behind timed-event coding is to capture behavior precisely as it unfolds, retaining its chronological order. A parsing rule can be to assign a new code every time a new team member speaks. If a single turn of talk includes several separate statements (e.g., a group member raises a problem and then asks a question), group researchers often decide to unitize more fine-grained thought or sense units (e.g., Bales, 1950; Keyton, 2018). Following this unitizing rule, a new code is assigned to every complete thought which is typically a phrase or single sentence but can also be a speech fragment (e.g., “Okay”).

In an interval-sampling plan, unitizing is based on pre-defined specific time intervals (e.g., Waller et al., 2004; Waller & Kaplan, 2018). For example, coders may assign a new code every 10 seconds (see Waller & Kaplan, 2018). Which unitizing approach researchers choose largely depends on the coding scheme that they select for their specific research project, and the constraints of the research environment. Some published coding schemes have specific unitizing rules and are tied to event sampling (e.g., Schermuly & Scholl, 2012; Keyton, 2004) while other published schemes are tied to an interval sampling plan (e.g., Waller et al., 2004; Waller & Kaplan, 2018). Both event-sampling and interval-sampling plans can be carried out in CAT. When using interval-sampling, CAT automatically reminds observers to log a new activity through a visual *shake* of the recording surface.

### **Immediacy: Live Versus Post-Hoc Observation**

Depending on the research specific design decisions, researchers can collect data on team interactions and their process dynamics either in real time using live coding (e.g., Farh & Chen, 2018; Manser et al., 2008; Schermuly & Scholl, 2012) or via audio or video recordings (e.g.,

Kaplan & Waller, 2016; Lehmann-Willenbrock & Allen, 2014). Which approach is preferred or suitable depends on a variety of factors. In field research, the particular study context is especially vital as it typically affording constraints. Within group and meeting settings, it often very difficult to record all participants in a way that their front (and face) is visible for the camera. In these cases, researchers either have to use a technically complicated arrangement of multi-camera videos that have to be synchronized, they have to ask participants to sit in a pattern that enhances visibility for the video recording (which is often unnatural and can disturb important group processes), or they have to accept that many participants are filmed with their back turned to the camera. In our particular example, and as described above, gathering audio or video recordings was only possible after a certain level of trust was built, and we therefore had to rely on live coding during the initial stages of data collection. Yet, video recordings might not be possible even when sufficient trust between the research team and the study participants has been built due to ethical, logistical, political, or legal reasons. Questions surrounding data privacy and data storage are especially important, and, as a result, many organizations are reluctant to agree to recordings. Likewise, video recording can at times not be possible simply due to logistical or budgetary reasons.

Today, data collection using videos (or detailed transcripts) is considered the highest standard in terms of data quality. One main advantage of coding from video recordings is that even subtleties in the interaction can be captured. By being able to rewind the recording (or playing it a slower speed), a trained coder can even capture parallel behavior or very minute behavior that is otherwise fast and fleeting (e.g., a smile). Likewise, recordings allow the behavior of team members to be captured at different levels (e.g., verbally and nonverbally). Furthermore, it is possible to evaluate the same recording by different observers and in different

passes, which plays an important part in establishing interrater reliability and validity of the coding. Furthermore, as the raw behavior becomes permanently accessible, researchers can refine their coding approach over time or add additional coding categories as more data is gathered or new knowledge arises.

Despite these clear benefits, working with video recordings is somewhat of a luxury in applied field research due to the reasons outlined above. This means that while video recordings might be preferable from most scientific and research perspectives, they can at times be not possible and—in such cases—alternative options need to be considered.

Accordingly, live observation is not uncommon in team research (e.g., Farh & Chen, 2018; Liu et al., 2019; Seelandt et al., 2018). For example, Manser and colleagues (2008) observed team coordination processes in medical teams during cardiac anesthesia. They had observers live-code a total of 36 different activities and reported good reliability. Another study by Seelandt et al. (2018) specifically compared live coding with video-based coding during medical team debriefs. Their coding system comprised 47 different communication behaviors. Findings showed that interrater reliabilities for live coding were equivalent to video coding (and in some cases even better). Schermuly and Scholl (2012) also compared live versus video coding using a coding scheme with five codes (and two rating scales). They reported satisfactory reliability for the real-time application and slightly higher reliabilities for the video-based application.

These examples highlight that live coding is feasible and can result in high quality data. When implementing live observation, researchers should keep in mind that the feasibility (and quality) of real-time coding is dependent on multiple factors. These factors involve the complexity of the coding scheme (in terms of both number and granularity of codes), the chosen

unitizing approach, the level of coder training, the team size, and the complexity of the team situation. For example, a rather calm expert panel meeting (as in our study) is likely easier to code than a hectic team emergency where multiple team members talk at once and team members physically interact with one another and various equipment and materials.

### **Software Options**

In terms of software options, researchers can choose between a range of different tools that will help them to collect, organize, and analyze observational data in a systematic way (for a detailed overview see Glüer, 2018). These available software options have developed from different research traditions such as developmental psychology (INTERACT), biology and animal research (Observer XT), psycholinguistics (ELAN), and qualitative research (e.g., MAXQDA, Nvivo, or AtlasTi). In selecting an appropriate software, there are a number of criteria to consider. We compiled a list of criteria that highlight some of the pros and cons of the existing options (building on recent work by Glüer, 2018). We acknowledge that this list may not be comprehensive. Furthermore, we constrained our list to software options that process video/audio data (e.g., INTERACT, Observer XT, Videograph, and ELAN) and excluded software that has been mostly applied within a qualitative research tradition and/or is mainly used to process text, documents, or pictures (e.g., MaxQDA, NVivo, Atlasti, or F4/F5).

Table 1 lists the available software options that fall in this domain. Among the criteria that we considered important, we noted (a) the costs for purchasing/using these software options, (b) their usability, (c) the required operating system, (d) options for coding schemes (e.g., whether researchers can customize coding schemes), (e) time-frame precisions of coded data, and (f) the possibility of real-time applications. We hope that this comparison will give readers a roadmap and decision point whether to use CAT or to choose another available option. Overall,

this comparison shows that CAT has certain strengths in terms of saving costs, its flexibility (operates in a web browser), extensive coding options, time accuracy, and real-time coding applications. In comparison, the other software options have their strengths when working with video data as they allow to play multiple videos at once and help in synchronization.

### **Brief Overview of the CAT Software**

The CAT software is easily accessible using a web-browser such as Firefox or Chrome. Data in CAT is secured via database security rules and a firebase authentication process. CAT is hosted by the *Centre for Transformative Work Design* at Curtin University, Western Australia, and can be accessed by online: <https://cat.ctwd.com.au>. Thus, CAT does not require a software package to be installed and there is no need to use separately stored files or external programs. The software is free if used for research purposes and can be used with a laptop, tablet, or smartphone.

Researchers interested in using CAT need to first create a user profile to sign in. They can then join an existing project (i.e., by receiving a weblink with a key for a specific “measure”) or create a new coding measure for their own project. Researcher who create a project can share it with other collaborators and assign different levels of authorization (e.g., permission to edit/change the coding measure, permission to access data). The observational data that are collected using CAT are saved in a tab-formulated matrix data file on a NoSQL cloud database cloud-server<sup>2</sup>. If users have no access to the internet during their data collection, they can download an offline version. As mentioned earlier, CAT software can be used for both live coding and for working with video or audio recordings. The software supports MP3 and MP4 format. When using the live-coding modus, CAT generates a time-stamped data file of the coded team activities. Timestamps indicate the actual time at which the team was observed and a



behavior was coded (e.g., asking a question, 19/01/2018, start time: 15:21:26; end time: 15:21:40). When coding from video/audio files, timestamps refer to time in the media file (e.g., asking a question, 19/01/2018, media file event start: 0:00:19.8; media file event end: 0:00:33.3).

In the following, we will describe four core features of CAT and provide examples from the research context in which we have developed and used the tool. First, we explain how researchers can create (or adapt) observational measures and integrate them in CAT to collect team process data. Second, in terms of data quality, we explain how CAT provides immediate feedback on interrater reliability. Third, CAT allows for visualizations of the coded team interaction data, which can be used to provide teams with immediate feedback. Fourth, we explain the export function that allows exporting data and using it for more advanced statistical analyses. Table 2 provides an overview and summarizes the main purpose of each feature (see column two), links them to examples from our research demonstration (see column three), and shows which questions can be addressed by each feature (see column four). Furthermore, Table 2 specifies the basic requirements necessary to use each feature.

### **First Core Feature: Creation of Measures**

In a first step, users either create their own coding measures (by using the *Measure* feature) and/or use an existing coding measure. When researchers create a new measure they have to give it a label. For example, they might label the measure “Interaction Process Analysis” when working with the IPA system (Bales, 1950, for a recent application of this scheme see Keyton & Beck, 2009). The codes from the scheme can then be organized using different *classes* that host codes with semantic similarities. For example, the codes from the IPA system can be organized into two larger classes, socio-emotional communication versus task-related communication, each comprising six fine-grained codes (examples for socio-emotional

communication are *showing solidarity* and *showing tension*; examples for task-orientation are *giving suggestions* and *asking for opinions*; Keyton & Beck, 2009). There are no restrictions regarding the number of codes that can be added to a measure. When adding new codes to a coding scheme, researchers can also add descriptions for each code which provide additional information (e.g., the code *showing solidarity* can receive a description such as “any act that shows positive feelings toward another person”). Furthermore, codes can be allocated a customized color which helps to cluster codes that are conceptually related.

Once the researcher has finalized a measure, it can be shared with other research collaborators using the share function (see Appendix B for an example). Likewise, collaborators can be added who would like to use the measure in another project. The share function generates a web link with a unique cooperation code that can be easily distributed via email to give others access to the coding scheme. This point was particularly important to us in the development of the tool. We wanted to create an online platform that could easily be used for international collaboration. Such a procedure is already standard for more classic survey research (e.g., using platforms like Qualtrics), but we were not aware of any similar online environment for behavioral research.

### **Our Example**

We developed a coding scheme to collect live observation data during the expert team meetings. Our goal was to create a group process system specifically focused on capturing knowledge sharing during problem solving in groups (Huang & Cummings, 2011; Kostopoulos & Bozionelos, 2011). Building on team research in knowledge-intensive teams and learning in organizations (Huang & Cummings, 2011; March, 1991), we were particularly interested in the interplay of exploratory knowledge (i.e., experimentation with new alternatives) and exploitative

knowledge (i.e., refinement and extension of existing knowledge) and their impact on team effectiveness. While previous research has studied this team phenomenon in the context of product innovation using social network analysis (Huang & Cummings, 2011), in military action teams (Knight, 2015), and by using more static/cross-sectional survey approaches (Kostopoulos & Bozionelos, 2011), we were interested in broadening the knowledge of how teams balance these dynamic team knowledge processes over time.

We labeled the measure *team communication analysis*. Next, we created a class that comprised a set of six mutually exclusive *functional codes*. These codes were based on a review of the literature on team exploration and exploitation (e.g., Huang & Cummings, 2011; Kostopoulos & Bozionelos, 2011; Knight, 2015) as well as previous research pertaining to problem solving in complex situations (e.g., Kanki et al., 1989; Stachowski et al., 2009; Uitdewilligen & Waller, 2018). Additional information on all codes, including sample statements, can be found in Table 3. We added a seventh code labeled *other communication* to make the class of behaviors exhaustive (Bakeman & Quera, 2011).

*Knowledge exploitation* was coded when team members utilized existing knowledge, which was indicated by expressing high knowledge certainty (e.g., verbalizing and describing the phenotype of the patient, reviewing the results of past diagnostic tests, and presenting knowledge about the case). An example was “There was some reduced white matter on it. It certainly looks like it was present and then with time really resolved.” *Knowledge exploration* was coded when a team member engaged in thought experiments, searched for ideas (e.g. “At that stage you don’t know if you’re dealing with the defect or really the B12 deficiency.”), or generated new ideas that were characterized by high levels of knowledge uncertainty or a need for feedback (“This is what I’ve seen in clinic and please, please help. I really don’t know what’s going on.”).

Knowledge exploration also included hypothesizing (“It could be some mechanism in terms of being able transport the B12.”) or presenting diagnostic ideas that were not confirmed yet (e.g., “We’ve hunted, and hunted, and hunted around all of those development pathways to try and find what is going on.”). *Inquiry* was coded when team members requested further information (mostly about the patient) or asked question about an analysis (e.g., “Was the brother tested ever?”). *Answer* was coded when team members answered a previous question by supplying additional information beyond acknowledgment (e.g., “Yeah, I think the brother has always had clearly normal markers.”). *Moving forward* was coded when team members made procedural suggestions (e.g., “We are going to hear today from the expert from dermatology.”) or recommended action steps and tasks to be carried out (e.g., “I think we have some steps to move forward . . .”). *Psychological safety behavior* was coded when team members made remarks that helped to build a climate of psychological safety within the meeting (e.g., “Please feel free to comment, there are no stupid comments.”) or that served as a protective mechanism for potential criticism (e.g., “Dysmorphism is in the eye of the beholder, but how I would describe this is that she has a large mouth, a relatively full lower lip [clinician gives a detailed description of facial features of the patient]”).

The first author shared the digital measure with two trained coders allowing them to access the project in CAT with their own laptop or tablet. All coders had a background in Psychology and had been involved in the research context for multiple weeks. Coders received 10-hours of training which covered theoretical concepts, working with pre-coded transcripts (training samples), comparisons of discrepancies on a point-by-point basis, and discussions to resolve these issues. Since coders coded the team live, all codes received a visual icon that helped to quickly identify each code during the meetings (see Figure 1).

Based on recommendations from previous observational studies focusing on teams in uncertain and non-routine situations (Waller, 1999; Waller et al., 2004; Waller & Kaplan, 2018), we selected a 15 second interval sampling strategy. That is, coders were prompted via CAT every 15 seconds to “code the behavior that was most salient/dominant during the last 15 seconds”. In initial field trials, we tested a unitizing approach which required observers to unitize and code events, but we realized that this approach was too challenging for a live coding project. Following this, we piloted an interval-sampling plan using 10 seconds (and later 15 sec) and found that the 15 sec interval to be the best solution for a real-time application in our study context. This duration provided a sufficient temporal resolution to capture changes in behavioral activities based on our coding scheme and it also allowed observers sufficient time to make a coding decision.

**Validity of coding scheme.** In terms of validity, we followed recommendations from Seelandt (2018) and used several methods to assess the validity of our newly developed coding scheme and its use for real-time applications. The validity of a coding scheme can be supported by multiple indices, including face validity, content validity, and convergent/divergent validity. While it is often not possible to assess all validity types, researchers are encouraged to demonstrate at least one form of validity (Seelandt, 2018). In our case, in terms of face validity (i.e., do codes simply look as if they are measuring the concepts of interest), we developed detailed descriptions for each code and also integrated these in the CAT software. We have reproduced these descriptions in Table 2. In terms of content validity (i.e., the degree to which the construct reflects all dimensions of interest), we reviewed the relevant literature on knowledge exploration and exploitation in teams (Hirst et al., 2018; Huang & Cummings, 2011; Håkansson et al., 2016; Jørgensen & Becker, 2017; Knight, 2015; Kostopoulos & Bozionelos,

2011; Nemanich & Vera, 2009) which guided the operationalization of these codes. Furthermore, we interviewed 16 participants from the expert panel on their perceptions of the meeting process. Before starting with the systematic observations, we also carried out unstructured observations (for about one year) that we used to scope out what behaviors tend to occur in the expert panel meetings, and we reviewed existing measures and coding schemes in order to adapt them for our purposes (see Table 3, column “related concepts”).

In terms of convergent (and discriminant) validity, we encouraged coders before the meetings to take short notes about their observations in the form of free text comments. Our goal was to use these transcribed notes to assess the convergent validity with an automated linguistic coding approach. In terms of coder instructions, we instructed them to only make summary notes that were particularly salient in capturing a specific code (e.g., if a participant said: “These spinal images show that the kid is having problems in region X and we can also see this in the diagnostic test [X]”, observers could note “discussion of images” to indicate knowledge exploitation). We also explained to coders that we could use their notes for discrepancy discussions (after a live coding) as these notes would help the team of coders to remember a specific episode.

Field text notes were logged in the CAT system and time-stamped (for an example see Table 4, column “remark”). We extracted these field notes and analyzed them using the linguistic inquiry and word count (LIWC) software (Pennebaker et al., 2015). Conceptually, knowledge exploration requires high levels of cognitive processing and builds on identifying discrepancies (“ought,” “should”) and tentative suggestions (“maybe,” “perhaps,” “I guess”). In contrast, knowledge exploitation requires less cognitive processing, builds on existing knowledge (i.e., known causes of a disorder in our case), and is directed towards the past (i.e., reviewing existing

information). Hence, we expected field notes that were coded as *exploration* to show higher proportions of words within linguistic categories such as “cognitive processes”, “discrepancies”, and “tentative” when compared with field notes coded as exploitation. Conversely, field notes coded as *exploitation* should contain a higher proportion of words within the linguistic categories “cause” and “past”.

We extracted 282 events that contained transcribed field notes (e.g., “possibly looks quite similar”, “could be external causes”, “I’m not a specialist but . . .”). We compiled the notes for each code (exploration and exploitation) and for each meeting. The LIWC software then quantified the percentage of word categories by comparing each word to a list of words within an internal dictionary.

In terms of convergent validity, knowledge exploration activities showed higher levels of cognitive processes ( $M = 0.25$ ,  $SD = 0.04$ ), discrepancies ( $M = 0.06$ ,  $SD = 0.01$ ), and tentativeness ( $M = 0.11$ ,  $SD = 0.04$ ), when compared to exploitation statements (cognitive processes:  $M = 0.09$ ,  $SD = 0.09$ ,  $t(4.3) = 3.4$ ,  $p = 0.02$ ; discrepancies:  $M = 0$ ,  $SD = 0$ ,  $t(3) = 7.8$ ,  $p = 0.004$ ; tentativeness:  $M = 0.01$ ,  $SD = 0.02$ ,  $t(4.6) = 4.9$ ,  $p = 0.005$ ). Knowledge exploration showed higher descriptive (albeit non-significant) levels of future focus ( $M = 0.01$ ,  $SD = 0.02$ ) and insights ( $M = 0.9$ ,  $SD = 0.01$ ) in comparison to exploitation activities (future focus:  $M = 0.00$ ,  $SD = 0$ ,  $t(3) = 1.43$ ,  $p = 0.25$ ; insights:  $M = 0.05$ ,  $SD = 0.05$ ,  $t(3.2) = 1.7$ ,  $p = .17$ )

Conversely, knowledge exploitation showed descriptively higher (albeit non-significant) levels of causation ( $M = 0.02$ ,  $SD = 0.04$ ) and past focus ( $M = 0.07$ ,  $SD = 0.02$ ) in comparison to exploration (causation:  $M = 0.01$ ,  $SD = 0.01$ ,  $t(6) = -0.86$ ,  $p = 0.42$ ; past focus:  $M = 0.03$ ,  $SD = 0.04$ ,  $t(6) = -1.82$ ,  $p = 0.12$ ). Overall, these analyses based on automated text analyses gave further support for the validity of the human coded activities.

There are also other ways to further establish discriminant validity with CAT. For example, when researchers code speakers (and their roles) in combination with the behavioral code (e.g., exploration vs. exploitation), they can compare the results for different groups (e.g., physicians versus nurses). Based on hierarchies or job descriptions, they could hypothesize that team members from a certain group (e.g., nurses) show a certain behavior (e.g., exploitation) more often than members from another group (e.g., physicians), which could provide evidence for discriminative validity.

### **Second Core Feature: Assessing Interrater Reliability**

CAT provides feedback on interrater reliability for each single code of the observational measure. For example, interrater reliability calculations for a rather detailed coding scheme with twenty codes would result in a total of twenty parameters. A code-specific interrater reliability allows the researcher to inspect which particular codes in the coding scheme are harder/easier for coders to detect. There are multiple reasons for this: Some codes may have unclear definitions, other codes may be harder to distinguish from conceptually similar codes, and some codes may just be harder to observe than others because they require higher levels of observer inference (e.g., “asking a question” might be easier to recognize than “feelings of unease”).

The session-based interrater reliability feedback implemented in CAT allows researchers to detect such *trouble-making codes* and gives them a more focused approach to coder training. That is, by inspecting code-specific interrater reliability estimates, the researcher can decide which codes need more attention in subsequent coder discussions, whether codes need to be re-defined, or whether semantically similar codes should be combined altogether. We designed this feature because we believe that interrater reliability calculations should not be considered as an afterthought in observational research but rather at the very beginning. Moreover, session-based



interrater reliability feedback allows the researcher to monitor the quality of incoming data on a session-by-session basis. A session constitutes a longer observation period such as a one-hour workplace meeting.

Interrater reliability values reported in CAT are calculated based on intraclass correlations (ICCs; McGraw & Wong, 1996) following recommendations reported in Hallgren (2012). Users can request estimates as soon as two independent coders have double-coded a single session (i.e., a single team performance episode).

Two types of ICCs are calculated, a consistency based (also called relative) ICC(R) and an absolute ICC(A). Both ICCs are calculated based on a two-way mixed ANOVA model. The relative based ICC(R) provides feedback on whether two (or more) observers' frequency scores are similar in relative rank order. The absolute based ICC(A) is more conservative and indicates whether absolute values are similar between observers (Hallgren, 2012). Researchers can decide which ICC variant is most suitable for their specific research question. If they want to know if the frequencies for each code are similar in *absolute value*, then absolute ICC(A) should be inspected. If researchers want to know if frequencies for each code are similar in rank order, then consistency should be inspected. To illustrate these differences, consider the following example in which two coders made observations during five consecutive sessions. The frequency counts of code A assigned by Coder 1 are generally low (2, 3, 6, 7, 10) whereas Coder 2 assigned code A to a larger extent (8, 9, 12, 13, 16). Because the frequencies are perfectly ordered in rank, the relative ICC(R) in value in this example reaches a maximum ( $ICC(R) = 1.00$ ). In contrast, the ICC(A) variant pays attention to absolute agreement and would be rather low in this example ( $ICC(A) = .36$ ). IRR feedback provides a preliminary reliability analysis and we recommend to code multiple sessions (i.e., at least five, ideally ten sessions, see Bakeman & Quera, 2011).

### **Our Example**

Figure 2 shows interrater reliability feedback for a single double-coded panel expert meeting that lasted for 50 minutes. This meeting was attended by 10 clinicians and was coded live by two coders. The top bar shows ICC(R) values and the bottom bar shows the corresponding ICC(A) values. To provide researchers with an intuitive reading of these values, CAT integrates color-coded cut-off values (the colors are displayed on the website) proposed by Cicchetti [1994; below .40 = poor (red); .40–.59 = fair (orange); .60–.74 = good (yellow); and .75–1.00 = excellent (green)]. We focus on the absolute ICC(A) values to evaluate if frequency scores between observers are similar in absolute value. Interrater reliability for the two codes *knowledge exploration* and *knowledge exploitation* yielded excellent values for this meeting (ICC(A) = .90 and ICC(A) = .80, respectively). The codes *moving forward* (ICC(A) = .60) and *psychological safety behavior* (ICC(A) = .60) can be both classified as good. The frequency measure for the code *inquiry* showed a fair reliability (ICC(A) = .55). Figure 2 also shows that it was not possible to compute an ICC score for the code *answer* since there were not enough observations (i.e., the first observer only coded this code once, while the second observer did not log this code at all). This session-based feedback allowed us to identify codes that required additional coder training. We used this information for the discrepancy discussions with our observers. We specifically discussed what instance they coded as indicative of psychological safety, moving forward, and inquiry to enhance mutual understanding of the coding instrument.

### **Third Core Feature: Feedback and Data Visualization**

After data collection, CAT allows immediate feedback and visualization of the annotated team data. This feedback feature does not require exporting the data and/or using external

software to obtain graphics, thus saving time and resources. Below, we highlight two different visualization options that are integrated into the software.

### **Visualization of Temporal Team Dynamics**

Coded team behaviors exhibit variability over time as well as between different team members (Kozlowski, 2015, Leenders et al., 2016). To explore such dynamics in coded behavior, CAT provides a visualization of the temporal team interaction process. This is accomplished by using Gantt charts that show when a particular code has been assigned during a team episode. Thus, this feature can help to identify the timing of specific behaviors (e.g., “at what time does a team need to exhibit a certain behavior to be effective?”, cf., Waller, 1999) or specific segments of team activities (Ballard et al., 2008). For example, researchers could see at a glance whether different codes were assigned at the beginning of the interaction than at the end of the interaction (e.g., is there more knowledge exploration at the beginning of the team interaction than at the end?).

Ballard et al. (2008) noted that distinct segments of the team process typically cannot be determined simply by dividing the performance episode in evenly distributed time intervals. That is, teams may engage in qualitatively different team episodes that vary in length and that can often not be determined a priori (Uitdewilligen & Waller, 2018). Illustrative research questions include: What episodes can be identified during multidisciplinary team meetings? How long should teams spend in each episode to be successful? What are breakpoints that determine transitions? or Is frequency or the specific timing (i.e., onset) of a behavior more important to explain team effectiveness? Which time points during the meeting are most important for knowledge exploration?

Second, in addition to behaviors, it is also possible to display when and for how long members of the team were active during a particular session. This setting can help to visually explore which members of the team showed peak activities throughout a particular session or performance episode (Koole et al., 2017). Illustrative research questions include: When are specific team members leading the discussion?, and Are there segments in the episode that involve increased turn-taking by all team members? This visualization is therefore particularly suitable for exploratory research questions and to build better theories of team adaptation.

Beyond exploratory research stages, visualizations also allow investigating specific team process phenomena. For example, van Oortmerssen and colleagues (2015) conducted a qualitative analysis of creative processes during governance board meetings and introduced the concept of interaction flow which they defined as “an optimal, intensified, and synergetic mode of the conversational interaction within a small group” (p. 522). They proposed that markers for interaction flows were intensified turn-taking dynamics among participants (i.e., short turns, wide distribution, and overlapping speaker times; van Oortmerssen et al., 2015). The visualization in CAT could help to identify these markers within the data stream.

### ***Our Example***

Figure 3 displays the temporal dynamics during one expert panel discussion. The behavioral codes of our coding measure are displayed on the *y*-axis, while the *x*-axis shows the temporal progression of the meeting. The plot shows when the codes occurred throughout the team meeting.

A visual exploration of these team interactions can be necessary as teams sometimes move in spurts such that coded behaviors are not evenly distributed across a performance episode (Ballard et al., 2008; Marks et al., 2001; Uitdewilligen & Waller, 2018). The phenomenon of

unevenly distributed communication has also been labeled as *team burstiness*, which describes the extent to which team members concentrate high levels of communication in a short period of time (Riedl & Woolley, 2017). The Gantt chart indicates that the team in our example mainly focused on knowledge exploitation (i.e., sharing what they already know about the patient and results of current genetic tests) during the beginning of the meeting (the first 7 min). Afterwards, Figure 3 indicates a dual process of knowledge exploration (i.e., participants expressing their ideas on what might be going on with the child) and exploitation (i.e., explanations of specific genetic tests or their current understanding of a particular rare disorder that shows similarities with the case) with a stronger focus on knowledge exploration towards the end of the meeting. There were also regular instances during the meeting that indicated that the team was moving forward. The final part of the meeting was mainly focused on answering questions as the team leader suggested to go around the room to collect different ideas. It is also noteworthy that around the midpoint of the meeting there were instances of psychological safety behaviors. These behaviors were placed to encourage experimentation and stimulate knowledge exploration. Overall, the temporal visualization can help to identify natural breakpoints in the process (cf., Uitdewilligen & Waller, 2018) and can provide a starting point to build theory around the nature of team process dynamics in a specific team performance context (in this case: diagnostic meetings between experts).

### **Team-Level Feedback**

Data on team processes can also be collected to carry out comparisons between different teams. That is, the focus of the research is on how variation in one team process explains variance in one or multiple team outcomes (e.g., Schmutz et al., 2015). CAT provides a team measurement function that summarizes the quantity of a team process measure for a single

session. This is visualized using a pie chart. This visualization helps to explore questions like “Which behaviors were most (or least) frequent during the session?” or “Do some teams show higher levels of certain team behaviors than others?” When combined with a team effectiveness measure, this also allows to answer questions like “Can we predict team performance based on variations of team communication behaviors?”

### ***Our Example***

Figure 4 displays a pie chart which provides a summary for each code of the team communication measure. Based on this meeting summary, we can see that the team showed rather balanced levels of knowledge exploitation (about 38%; e.g., reviewing shared information about the case such as results about genetic tests, and stressing distinct phenotypical observations that are clearly linked with known genetic dysfunctions) and knowledge exploration behaviors (about 44%; e.g., hypothesizing about potential disorders that could explain some puzzling observations).

Research on innovation teams suggest that both team exploratory learning and team exploitative learning are important predictors of team performance (Kostopoulos & Bozionelos, 2011). Hence, collecting data from multiple meetings (or teams) in knowledge-intensive multi-disciplinary expert teams allows us to substantiate and expand these finding within novel organizational contexts. Furthermore, this feature allows to carry out team feedback interventions that help teams to reflect on their learning processes during these meetings. This also enables research designs in which these assessments may stimulate positive changes in team learning processes (e.g., by comparing teams that receive feedback to teams that receive no feedback).

### **Fourth Core Feature: Export of Raw Data and Analyses**

CAT saves all collected data into a two-dimensional data file (i.e., columns, rows) that can be easily exported for use in other statistical programs such as SPSS or R. This allows researchers to process the data for more complex statistical analyses (e.g., regression, ANOVA, pattern analysis). Once the data have been collected, two different output files, CSV and Excel, can be generated.

Collected data are organized into *sessions*. In the context of team and small group research, a session usually refers to a team performance episode. Typical examples for a session are a medical operation (Kolbe et al., 2014), a flight of an aviation team (Lei et al., 2016), or an emergency management situation (Uitdewilligen & Waller, 2018). Depending on these contexts, these sessions can be labeled accordingly (e.g., when observing multiple meetings, a session name could be “Sales pitch meeting: 10-10-2018”).

### **Our Example**

Table 4 shows the multi-level structure of the output file for our particular example. Output files generated by CAT contain a column that indicates which session the annotation belongs to (here: Meetings for “Patient X23” and for “Patient M4”) and a column that shows which observer coded the data (here: PL and MH). These are examples of higher-order variables because they show no variation for a particular session. Naturally, each session will have multiple observations of coded activities. These are written into separate columns and are examples of lower-level variables. In line with the particular coding scheme or measure that was specified beforehand, CAT provides a column for the behavioral code that was assigned to each event (*Selected Category*). If the coders made use of additional field notes, these are listed in a separate column (*Remark*). For example, Table 4 shows that observer MH logged three events in

the session for patient X23. One event was coded as *exploitation*, one event was coded as *exploration*, and one event was coded as *other communication*. Finally, CAT creates an index for all events and each observer (*Interval index*) and saves detailed information on the specific onset and offset times of each coded event. In Table 4, the time format follows the 24-hour time clock notation (HH:MM:SS) because observations were coded live. The duration for each coded event is shown in a standard time format with minutes, seconds, and milliseconds.

In sum, export data files are multi-level two-dimensional files containing all behavioral codes generated by the observer. Data are displayed in a timed-event sequential order. When researchers work with video-recorded files, the data file also includes labels for the respective media files, creator country, file location of the specific video used, and the date and time the video was coded.

In terms of analyses, the structured output matrix data file allows researchers to answer different types of research questions. In our research, we are interested in exploring the following questions: What is the temporal pattern of knowledge exploration in the context of teams working under extreme uncertainty? Furthermore, we want to know whether effective meetings are characterized by different exploration dynamics when compared to ineffective meetings. In other words, what characterizes the temporal pattern of knowledge exploration for effective versus ineffective expert meetings?

Existing research on exploration dynamics in teams suggests that team exploration shows time-dependent patterns over the course of a performance episode, that is, exploration shows an initial positive growth, peaks around the midpoint, and then slowly declines in the second half of a performance episode (Knight, 2015). Furthermore, this research suggests that higher levels of early exploration and a steeper decline in exploration during the second half of a performance



episode are associated with better team performance. While our research focuses on one-hour performance episodes, specifically expert panel meetings, within a knowledge-working context of health experts, previous research on exploration dynamics has taken place in very different contexts (i.e., military/action teams) and over significantly different performance episodes (i.e., 16 weeks, cf., Knight, 2015). Nonetheless, extrapolating from this earlier research, we could hypothesize that knowledge exploration within the expert meetings shows a curvilinear slope over time (H1) and that effective expert meetings have different temporal exploration patterns than ineffective meetings (H2).

To test these hypotheses, we can use CAT to collect data from multiple meetings using the coding scheme outlined above. To assess meeting effectiveness, we could use a multi-item measure that participants complete at the end of each meeting. For example, we could ask participants to report how effective each meeting was in terms of generating novel insights/eureka moments, in narrowing down the number of possible diagnoses for the patient, and in improving the patient management.

After coding multiple sessions, the exported CAT output files can be used to calculate repeated measures for exploration within each meeting. Specifically, we could compute count scores for exploration by dividing meetings into four equal quarters (i.e., four repeated measurements). To test if exploration in meetings has a curvilinear slope over time across all meetings (H1), and whether these patterns are associated with meeting effectiveness (H2), we could calculate a repeated measurement ANCOVA with exploration as the focal outcome. In the model, time (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup> meeting quarter) would be the predictor and meeting effectiveness the covariate. If we were to find a main effect of quadratic time on knowledge exploration, we could support H1, which proposed a curvilinear slope of exploration. The interaction between

time and meeting effectiveness would allow us to test the second hypothesis, which stated that more effective meetings are characterized by different exploration process dynamics than less effective meetings. To understand the nature of the interaction effect on knowledge exploration, we could plot the average slope for high versus for low effective meetings. The interaction plot should show that more effective meetings are characterized by high levels of exploration during the first half of the meeting and a steep decrease in knowledge exploration later in the meeting (i.e., in the second meeting half). In contrast, meetings with lower effectiveness should display a different pattern (with high levels of early exploration and lower levels of late exploration). Overall, these analyses would allow us to generate insights into the role of time in groups and to better take into account the dynamic nature of team processes.

### **Practical Applications of the CAT Software**

We certainly see practical applications in using CAT, in particular, for collecting high resolution data to better understand team dynamics (Klonek et al., 2019; Kolbe & Boos, 2019). In a recent review, the authors have identified multiple industry and applied field contexts that may benefit from this approach (Klonek et al., 2019), such as, in high fidelity contexts (e.g., cockpit crews, nuclear power plant crews) or teams in extreme environments (e.g., aeronautics space teams, emergency response teams, polar expeditions). Moreover, Kolbe and Boos (2019) have proposed multiple examples for studying fine-grained dynamics within healthcare. The visualizations of the coded team data could be used to provide the teams with direct feedback about their social dynamics.

CAT can also be used to complement team training (Hughes et al., 2016) or team development interventions (e.g., Shuffler et al., 2011). Professional behavior-based feedback can help to stimulate important reflection processes in the team and provide a starting point for team

learning. The feedback function implemented in CAT allows to provide timely feedback on team dynamics while interactions from performance episodes are still fresh in the team's mind. It also allows the team to understand how specific team activities are shifting over time. For example, in a team training based on CAT, the visualizations could be used to highlight that the team has engaged in planning behavior too late (or too early) during a performance episode. Likewise, we see potential for using CAT for more standardized team training such as team training for medical emergency teams. Such teams often have to follow standardized procedures and can highly benefit from specific behavior-based feedback (e.g., Kolbe et al., 2014). For example, teams could be observed and coded during life-threatening cardiac arrest emergencies. Feedback based on CAT could be used to point out when the team needs to engage in closed-looped communication which showed to be particularly helpful for coordination when the task at hand requires clear steps to follow and which is critical for patient safety (e.g., Schmutz et al., 2015). In such high risk environments involving actual patients, organizations might be more open to methods that do not necessarily require video-recordings and thus the live coding option implemented in CAT might be especially useful for highly quality debriefs.

### **Limitations of the CAT Software**

Despite various benefits of using CAT for data collection and analyses, the tool is not without limitations. First, CAT is particularly well suited for data collection contexts in which researchers use either video- or live observations. The tool is not necessarily superior to situations in which researcher have access to a transcribed dataset, such as existing virtual team communication chat logs (e.g., Schechter et al., 2017). However, when researchers are unsure whether to transcribe team communication (due to better triangulation of concepts), they should weigh the additional costs and efforts that are involved in creating high quality transcripts.

Furthermore, when researchers have the choice between live versus video/audio-recorded team, we would encourage them to use (whenever possible) the video-based approach – particularly in instances when coder fatigue is increasingly likely. In research projects that do require live coding over multiple hours, future research with CAT could allow to test if reliability changes as a function of coding time. This would allow to get a better understanding of how coder fatigue potentially impacts the quality of live coded data.

We strongly encourage researchers to continue to utilize piloting approaches and other methods available to ensure the quality of the observational data being collected. Coding a media-file with CAT allows researcher to stop, replay, and recode the team interactions, which should result in better triangulation. Finally, we need more future studies that compare the reliability of coding schemes in a live versus media-based and CAT could facilitate this type of research.

Second, kappa calculations (e.g., Cohen's kappa; Cohen, 1960) are not yet available in the software. Whereas ICC values provide insights into the reliability of frequency measures, Kappa is a point-by-point agreement metric for categorical data. It can thus answer whether two or more observers have used the same codes in the same order. We intend to add this functionality at a later point in time.

Third, CAT was developed as a tool for quantitative analysis of team interactions (cf. Keyton, 2018; Waller & Kaplan, 2018). While the method of interaction analysis shares certain overlap with qualitative and mixed-method type research (e.g., using coding schemes to transform qualitative data into quantitative data; Gibson, 2017; see also Keyton, 2018), its main focus is quantitative and it is often limited to a reduced number of codes to answer specific questions. From a qualitative lens, many things can happen within 15 seconds and we

acknowledge that some design decisions in this quantitative research tradition (e.g., assigning the most dominant code to a 15 sec interval) can significantly reduce data richness.

### **Conclusion**

The CAT software introduced in this article offers team researchers the opportunity to index dynamic team phenomena within real organizational contexts. Specifically, team researchers using CAT can collect large amounts of time-stamped data which enables research on temporal process dynamics (Klonek et al., 2019). This way, researchers can test temporal theories about team constructs and advance our knowledge on time-dependent phenomena.

CAT was designed to align with constraints and opportunities of team research *in the wild*. The software has a high level of mobility; that is, researchers with a tablet can use the web browser or offline version, and share their measures and data across research sites and laboratories. We hope to have sparked the interest of researchers in this field and ultimately to promote better opportunities for collaboration.

We see various application areas in which CAT might prove helpful, spanning both research and practice. We think it is a promising time for more dynamic team process research and are particularly happy to see increased interest in systematic observation research.

### References

- Bakeman, R., & Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press.
- Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Addison-Wesley.
- Ballard, D. I., Tschan, F., & Waller, M. J. (2008). All in the timing: Considering time at multiple stages of group research. *Small Group Research, 39*, 328–351.  
<https://doi.org/10.1177/1046496408317036>
- Barto, D., Bird, C. W., Hamilton, D. A., & Fink, B. C. (2016). The Simple Video Coder: A free tool for efficiently coding social video data. *Behavior Research Methods, 49*, 1563–1568.  
<https://doi.org/10.3758/s13428-016-0787-0>
- Baynam, G., Broley, S., Bauskis, A., Pachter, N., McKenzie, F., Townshend, S., ... & Schofield, L. (2017). Initiating an undiagnosed diseases program in the Western Australian public health system. *Orphanet Journal of Rare Diseases, 12*, 83.  
<https://doi.org/10.1186/s13023-017-0619-z>.
- Baynam, G., Pachter, N., McKenzie, F., Townshend, S., Slee, J., Kiraly-Borri, C., ... & Verhoef, H. (2016). The rare and undiagnosed diseases diagnostic service—application of massively parallel sequencing in a state-wide clinical service. *Orphanet Journal of Rare Diseases, 11*, 77. <https://doi.org/10.1186/s13023-016-0462-7>
- Brauner, E., Boos, M., & Kolbe, M. (Eds.). (2018). *The Cambridge handbook of group interaction analysis*. Cambridge University Press.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>

- Cronin, M. A., Weingart, L. R., & Todorova, G. (2011). Dynamics in groups: Are we there yet? *Academy of Management Annals*, *5*, 571–612.  
<https://doi.org/10.1080/19416520.2011.590297>
- Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*(1), 1-19. <https://doi.org/10.1007/s10869-013-9308-7>
- Edmondson, A. C. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, *40*, 1419–1452.  
<https://doi.org/10.1111/1467-6486.00386>
- Eisen, L. A., & Savel, R. H. (2009). What went right: Lessons for the intensivist from the crew of US Airways Flight 1549. *Chest*, *136*, 910–917. <https://doi.org/10.1378/chest.09-0377>.
- Farh, C. I., & Chen, G. (2018). Leadership and member voice in action teams: Test of a dynamic phase model. *Journal of Applied Psychology*, *103*, 97–110.  
<https://doi.org/10.1037/apl0000256>
- Glüer, M. (2018). Software for coding and analysing interaction processes. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis*. (pp. 245–273). Cambridge University Press.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34.  
<https://doi.org/10.20982/tqmp.08.1.p023>
- Håkonsson, D. D., Eskildsen, J. K., Argote, L., Mønster, D., Burton, R. M., & Obel, B. (2016). Exploration versus exploitation: Emotions and performance as antecedents and consequences of team decisions. *Strategic Management Journal*, *37*, 985–1001.  
<https://doi.org/10.1002/smj.2380>

- Hirst, G., van Knippenberg, D., Zhou, Q., Zhu, C. J., & Tsai, P. C. F. (2018). Exploitation and exploration climates' influence on performance and creativity: Diminishing returns as function of self-efficacy. *Journal of Management*, *44*, 870–891.  
<https://doi.org/10.1177/0149206315596814>
- Huang, S., & Cummings, J. N. (2011). When critical knowledge is most critical: Centralization in knowledge-intensive teams. *Small Group Research*, *42*, 669–699.  
<https://doi.org/10.1177/1046496411410073>
- Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. *Journal of Applied Psychology*, *101*, 1266–1304.  
<https://doi.org/10.1037/apl0000120>
- Jørgensen, F., & Becker, K. (2017). The role of HRM in facilitating team ambidexterity. *Human Resource Management Journal*, *27*, 264–280. <https://doi.org/10.1111/1748-8583.12128>
- Kanki, B. G., Lozito, S., Foushee, H. C. (1989). Communication indices of crew coordination. *Aviation Space and Environmental Medicine*, *60*, 56–60.
- Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on team and organizational success. *Small Group Research*, *43*, 130–158.  
<https://doi.org/10.1177/1046496411429599>
- Kennedy, D. M., & McComb, S. A. (2014). When teams shift among processes: Insights from simulation and optimization. *Journal of Applied Psychology*, *99*, 784–815.  
<https://doi.org/10.1037/a0037339>
- Keyton, J. (2018). Interaction analysis: An introduction. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis*. (pp. 3–19). Cambridge University Press.



- Keyton, J., & Beck, S. J. (2009). The influential role of relational messages in group interaction. *Group Dynamics: Theory, Research, and Practice, 13*, 14–30.  
<https://doi.org/10.1037/a0013495>
- Klonek, F.E., Gerpott, F., Lehmann-Willenbrock, N., & Parker, S. (2019). Time to go wild: How to conceptualize and measure process dynamics in real teams with high resolution? *Organizational Psychology Review*. Advance online publication.  
<https://doi.org/10.1177/2041386619886674>
- Klonek, F. E., Quera, V., & Kauffeld, S. (2015). Coding interactions in motivational interviewing with computer-software: What are the advantages for process researchers? *Computers in Human Behavior, 44*, 284–292. <https://doi.org/10.1016/j.chb.2014.10.034>
- Knight, A. P. (2015). Mood at the midpoint: Affect and change in exploratory search over time in teams that face a deadline. *Organization Science, 26*, 99–118.  
<https://doi.org/10.1287/orsc.2013.0866>
- Kolbe, M., & Boos, M. (2019). Laborious but elaborate: The benefits of really studying team dynamics. *Frontiers in Psychology, 10*, 1478. <https://doi.org/10.3389/fpsyg.2019.01478>
- Kolbe, M., Grote, G., Waller, M. J., Wacker, J., Grande, B., Burtscher, M. J., & Spahn, D. R. (2014). Monitoring and talking to the room: Autochthonous coordination patterns in team interaction and performance. *Journal of Applied Psychology, 99*, 1254–1267.  
<https://doi.org/10.1037/a0037877>
- Koole, T., van Burgsteden, L., Harms, P., van Diemen, C. C., & van Langen, I. M. (2017). Participation in interdisciplinary meetings on genetic diagnostics (NGS). *European Journal of Human Genetics, 25*, 1099–1105. <https://doi.org/10.1038/ejhg.2017.111>

- Kostopoulos, K. C., & Bozionelos, N. (2011). Team exploratory and exploitative learning: Psychological safety, task conflict, and team performance. *Group & Organization Management, 36*, 385–415. <https://doi.org/10.1177/1059601111405985>
- Kozlowski, S. W. (2015). Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review, 5*, 270–299. <https://doi.org/10.1177/2041386614533586>
- Kozlowski, S. W., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods, 16*, 581–615. <https://doi.org/10.1177/1094428113493119>
- Leenders, R. T. A., Contractor, N. S., & DeChurch, L. A. (2016). Once upon a time: Understanding team processes as relational event networks. *Organizational Psychology Review, 6*, 92–115. <https://doi.org/10.1177/2041386615578312>
- Lehmann-Willenbrock, N., & Allen, J. A. (2018). Modeling temporal interaction dynamics in organizational settings. *Journal of Business and Psychology, 33*, 325–344. <https://doi.org/10.1007/s10869-017-9506-9>
- Lei, Z., Waller, M. J., Hagen, J., & Kaplan, S. (2016). Team adaptiveness in dynamic contexts: Contextualizing the roles of interaction patterns and in-process planning. *Group & Organization Management, 41*, 491–525. <https://doi.org/10.1177/1059601115615246>
- Liu, Y., Vashdi, D. R., Cross, T., Bamberger, P., & Erez, A. (2019). Exploring the puzzle of civility: Whether and when team civil communication influences team members' role performance. *Human Relations*. Advance online publication. <https://doi.org/10.1177/0018726719830164>
- Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2018). A fitting approach to construct and measurement alignment: The role of big data in advancing dynamic

- theories. *Organizational Research Methods*, 21, 592–632.  
<https://doi.org/10.1177/1094428117728372>
- Maclin, O. H., & Maclin, M. K. (2005). Coding observational data: A software solution. *Behavior Research Methods*, 37, 224–231. <https://doi.org/10.3758/bf03192690>
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2, 71–87. <https://doi.org/10.1287/orsc.2.1.71>
- Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26, 356–376.  
<https://doi.org/10.2307/259182>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46. <https://doi.org/10.1037/1082-989x.1.1.30>
- Meinecke, A. L., Klonek, F. E., & Kauffeld, S. (2016). Using observational research methods to study voice and silence in organizations. *German Journal of Human Resource Management*, 30, 195–224. <https://doi.org/10.1177/2397002216649862>
- National Transportation Safety Board (NTSB) (2010). *Loss of thrust in both engines after encountering a flock of birds and subsequent ditching on the Hudson river, US Airways Flight 1549, Airbus A320-214, N106US, Weehawken, New Jersey, January 15, 2009* (Aircraft Accident Report NTSB/AAR-10/03). National Transportation Safety Board Website. <https://www.nts.gov/investigations/AccidentReports/Reports/AAR1003.pdf>
- Nemanich, L. A., & Vera, D. (2009). Transformational leadership and ambidexterity in the context of an acquisition. *The Leadership Quarterly*, 20, 19–33.  
<https://doi.org/10.1016/j.leaqua.2008.11.002>
- Noldus, L. P. J. J., Trienes, R. J. H., Hendriksen, A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer Video-Pro: New software for the collection, management, and presentation

- of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, & Computers*, 32, 197–206. <https://doi.org/10.3758/bf03200802>
- Oborn, E., & Dawson, S. (2010). Knowledge and practice in multidisciplinary teams: Struggle, accommodation and privilege. *Human Relations*, 63, 1835–1857. <https://doi.org/10.1177/0018726710371237>
- Pennebaker, J.W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.
- Riedl, C., & Woolle, A. W. (2017). Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance. *Academy of Management Discoveries*, 3, 382–403. <https://doi.org/10.5465/amd.2015.0097>
- Roe, R. A., Gockel, C., & Meyer, B. (2012). Time and change in teams: Where we are and where we are moving. *European Journal of Work and Organizational Psychology*, 21, 629–656. <https://doi.org/10.1080/1359432X.2012.729821>
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50, 540–547. <https://doi.org/10.1518/001872008X288457>
- Schechter, A., Pilny, A., Leung, A., Poole, M. S., & Contractor, N. (2017). Step by step: Capturing the dynamics of work team process through relational event sequences. *Journal of Organizational Behavior*, 39, 1163–1181. <https://doi.org/10.1002/job.2247>
- Schermuly, C. C., & Scholl, W. (2012). The Discussion Coding System (DCS)—A new instrument for analyzing communication processes. *Communication Methods and Measures*, 6, 12–40. <https://doi.org/10.1080/19312458.2011.651346>
- Schmutz, J., Hoffmann, F., Heimberg, E., & Manser, T. (2015). Effective coordination in

- medical emergency teams: The moderating role of task type. *European Journal of Work and Organizational Psychology*, 24, 761–776.  
<https://doi.org/10.1080/1359432X.2015.1018184>
- Schmutz, J. B., Lei, Z., Eppich, W. J., & Manser, T. (2018). Reflection in the heat of the moment: The role of in-action team reflexivity in health care emergency teams. *Journal of Organizational Behavior*, 3, 749–765. <https://doi.org/10.1002/job.2299>
- Seelandt, J. C. (2018). Quality control: Assessing reliability and validity. In E. Brauner, M. Boos, & M. Kolbe (Eds.), *The Cambridge handbook of group interaction analysis*. (pp. 227–244). Cambridge University Press.
- Seelandt, J. C., Grande, B., Kriech, S., & Kolbe, M. (2018). DE-CODE: a coding scheme for assessing debriefing interactions. *BMJ Simulation and Technology Enhanced Learning*, 4(2), 51-58. <http://dx.doi.org/10.1136/bmjstel-2017-000233>
- Seelandt, J. C., Grande, B., Kriech, S., & Kolbe, M. (2018). DE-CODE: A coding scheme for assessing debriefing interactions. *BMJ Simulation and Technology Enhanced Learning*, 4, 51-58. <https://doi.org/10.1136/bmjstel-2017-000233>
- Shuffler, M. L., Diaz Granados, D., & Salas, E. (2011). There's a science for that: Team development interventions in organizations. *Current Directions in Psychological Science*, 20, 365–372. <https://doi.org/10.1177/0963721411422054>
- Stachowski, A. A., Kaplan, S. A., & Waller, M. J. (2009). The benefits of flexible team interaction during crises. *Journal of Applied Psychology*, 94, 1536–1543.  
<https://doi.org/10.1037/a0016903>
- Uitdewilligen, S., Rico, R., & Waller, M. J. (2018). Fluid and stable: Dynamics of team action patterns and adaptive outcomes. *Journal of Organizational Behavior*, 39, 1113–1128.  
<https://doi.org/10.1002/job.2267>

- Uitdewilligen, S., & Waller, M. J. (2018). Information sharing and decision-making in multidisciplinary crisis management teams. *Journal of Organizational Behavior, 39*, 731–748. <https://doi.org/10.1002/job.2301>
- van Oortmerssen, L. A., van Woerkum, C. M., & Aarts, N. (2015). When interaction flows: an exploration of collective creative processes on a collaborative governance board. *Group & Organization Management, 40*, 500-528. <https://doi.org/10.1177/1059601114560586>
- Walker, C. E., Mahede, T., Davis, G., Miller, L. J., Girschik, J., Brameld, K., Wenxing, S., Rath, A., Ségolène, A., Zubrick, S., Baynam, G. S., Molster, C., Dawkins, H. J. S., & Weeramanthri, T.S. (2017). The collective impact of rare diseases in Western Australia: an estimate using a population-based cohort. *Genetics in Medicine, 19*(5), 546.
- Waller, M. J. (1999). The timing of adaptive group responses to nonroutine events. *Academy of Management Journal, 42*, 127–137. <https://doi.org/10.1016/j.jm.2003.07.001>
- Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science, 50*, 1534–1544. <https://doi.org/10.1287/mnsc.1040.0210>
- Waller, M. J., & Kaplan, S. A. (2018). Systematic behavioral observation for emergent team phenomena: Key considerations for quantitative video-based approaches. *Organizational Research Methods, 21*, 500–515. <https://doi.org/10.1177/1094428116647785>

#### Endnote

<sup>1</sup> We want to point out that readers should not confuse the term *group interaction* analysis with the analytical approach of testing moderations using multiple regression (Dawson, 2014) which is sometimes also labelled as *interaction analysis*. Group interaction analysis as referred to here is “a systematic research technique for reliably unitizing and coding naturally occurring interaction behaviors and making valid interpretations and inferences from those data to the context in which the observations occurred” (Keyton, 2018, p. 3)

<sup>2</sup> <https://firebase.google.com/docs/database/>

<sup>3</sup> A detailed demonstration for how to access CAT and use it are provided on the website:  
[cat.ctwd.au](http://cat.ctwd.au)

Table 1.

*Comparison of Existing Coding Tools for Collecting, Organizing, and Analyzing Team Interaction Data*

	CAT	INTERACT	Observer XT	Videograph	ELAN
Webpage	cat.ctwd.com.au	www.mangold-international.com/en	www.noldus.com	www.dervideograph.de	https://tla.mpi.nl/tools/tla-tools/elan/
Research tradition	Team/ Group dynamics, Organizational Behavior	Developmental Psychology	Animal Research	Teaching / Education	Psycholinguistics
Usability <sup>b</sup>	High*	High	Moderate	Moderate	Low
Operation system	Web-browser	Windows	Windows	Windows	Windows, MaxOS, Linux
Coding schemes/ options <sup>c</sup>	Extensive coding management (e.g., visual icons/code descriptions, mouse/double-click for coding, shortcuts for media-files)	Comprehensive coding management (e.g., programmable keyboard shortcuts)	Comprehensive coding management (e.g., programmable keyboard shortcuts)	Basic / restricted coding management (e.g., only ten programmable keyboard shortcuts, no letters; only 40 codes)	Basic / restricted coding management (e.g., coding by mouse or keyboard not possible)
Time precision	Milliseconds	Picture frames	Picture frames	Seconds	Milliseconds
Real-time coding & portable app	Included	Obansys (extra product)	Pocket Observer (extra product)	n.a.(only video)	n.a.(only video and audio)



Pros	-High flexibility (no extra license needed, browser application) - Live coding options - Data visualization - Easy data sharing	-Plays multiple videos simultaneously -Technical user support (with a license) -Comprehensive analysis and visualization options	-Plays (only) two videos simultaneously -Technical user support (with a license)	-Plays multiple videos simultaneously	-Plays multiple videos simultaneously -Free
Cons	-Plays only one video-file at a time - User support only through authors/developers	-Best to use for video/audio-coding -High costs - program-specific data files (.iact) which can only be opened with the software	-Best to use for video/audio-coding - High costs - program-specific data files (.odf) which can only be opened with the software	-Website and manual are in German only -Error-prone -Costs -User support only through authors	-Technical terminology from linguistics impedes usability -User support only through forum

*Note.* In terms of costs, CAT and ELAN are free programs; previous cost estimates for the other programs can be found in Glüer (2018) and Lehmann-Willenbrock and Allen (2018); ELAN = *Eudico Linguistic Annotator*; a = b = Usability ratings used from Glüer (2018) and Lehmann-Willenbrock and Allen (2018); c = ‘coding options’ information provided by Glüer (2018); \* = rating of CAT usability are based on subjective experiences from two of our authors who have had experiences working with CAT and INTERACT

Table 2.

*Core Features of the Communication Analysis Tool*

Feature	Purpose	Example	Question that can/should be addressed	Requirements
Creation of measure	<i>For the researcher:</i> Operationalization of one or multiple team phenomena	Table 3, Figure 1	Which team process variables are relevant to answer the particular research question at hand?	Access to an existing coding scheme or creation/adaptation of new coding scheme
Reliability of measure	<i>For the researcher:</i> Assessment of interrater reliability of the coded data	Figure 2	Is the team process measure reliable?	At least two (trained) observers/coders who independently analyse the same team during a critical performance episode (e.g., a meeting)
Visualization of team processes	<i>For the research participants:</i> Feedback to research participants, illustrating the value of the research, retention of participants, summary of team process dynamics, identification of aspects for an intervention	Figure 3, Figure 4	How do team processes vary across a performance episode? Do some team members show higher participation than others? Are there transition points in the performance episode that indicate a shift in team process dynamics?	One (trained) observer/coder who has access to a team during a critical performance episode
Export of data	<i>For the researcher:</i> Combination with other data sources (e.g., survey data or performance measures) for testing research hypotheses	Table 4	Which variables explain temporal variations in team processes? Do effective teams show different team process dynamics or patterns than ineffective teams?	Use of statistical programs (e.g., SPSS) to triangulate coded team process data with other data sources (e.g., team effectiveness ratings), collection of multiple teams/team episodes

Table 3.

*Coding Scheme used in Research Project on Multidisciplinary Expert Meetings in Health Care*












Code	Definition	Example	Related concepts
 — Knowledge exploitation	Knowledge-related contributions with high certainty; utilization and processing of existing knowledge: Reviewing and describing information which have a high level of certainty (e.g., phenotype descriptions, facts about the patient, previous clinical test results, knowledge about the case)	(1) In summary, we have two children who have combination of varying degrees of learning difficulty... (2) There was some reduced white matter on. It certainly looks like it was present and then with time really resolved.	<i>Fact sharing</i> (Uitdewilligen & Waller, 2018), <i>Reviewing and situation assessment</i> (Schmutz, Lei, Eppich, & Manser, 2018), <i>team exploitative learning</i> (Kostopoulos & Bozionelos, 2011)
 — Knowledge exploration	Knowledge-related contributions with a low degree of certainty (e.g., hypothesizing, presenting diagnostic ideas that are not confirmed yet, expressing of uncertainty, exploratory remarks; searching, experimentation, and developing new ideas)	(1) “This is what I've seen in clinic and please, please help. I really don't know what's going on.” (2) We've hunted, and hunted, and hunted around all of those development pathways to try and find. (3) I actually don't know. I don't know.	<i>Interpretation sharing</i> (Uitdewilligen & Waller, 2018), <i>exploration of solution space</i> (March, 1991), <i>team exploratory learning</i> (Kostopoulos & Bozionelos, 2011)
<b>?</b> — Inquiry and Questions	Request for information, statement, or analysis, (Cognitive activity with an interpersonal direction)	(1) It's the same girl? (2) Was the brother tested ever?	<i>Inquiry and Question</i> from Lei et al. (2016)
<b>→</b> —Answer	Supplying information beyond acknowledgment (Cognitive activity with interpersonal direction)	(1) It's the same girl. (2) Yeah, I think the brother has always had clearly normal markers.	<i>Answer</i> from Lei et al. (2016)
 —Action & Moving Forward	Suggestions for further procedure, recommendation for action	(1) I'm going rely a lot on [Name] today from the metabolic, B-12 perspective in particular. (2) I'll show you some of the pictures in a second. (3) So [Name], I think these are your ... ]directs the conversation to another participant]	<i>Positive procedural communication (procedural suggestion, prioritizing, task distribution), and action planning</i> (Kauffeld & Lehmann-Willenbrock, 2012)
 — Psychological Safety Behaviors	Remarks that help to increase the psychological safety in the meeting (e.g., remarks that one's opinion is not failsafe; genuine comments that encourage “stupid” questions)	“Dysmorphism is in the eye of the beholder, but how I would describe it...” “I might be completely wrong, but couldn't it be...” “There is no such thing as stupid comments...”	<i>Team psychological safety</i> (Edmondson, 2003)

Table 4.

*Structure of Exported Data File*

Session	Date	Observer	Interval Index	Selected Category	Remark	Event Start Time	Event End Time	Duration
Patient X23	19/03/2019	PL	4	 Exploitation	x was done...	10:13:09	10:13:23	00:14.123
Patient X23	19/03/2019	PL	5	 Exploration	'possibly looks quite similar'	10:13:23	10:13:37	00:13.887
Patient X23	19/03/2019	PL	6	 Exploitation		10:13:37	10:13:52	00:14.664
Patient X23	19/03/2019	PL	7	 Exploitation	knee ultrasound	10:13:52	10:14:07	00:15.056
Patient X23	19/03/2019	PL	8	Other communication		10:14:07	10:14:21	00:14.616
Patient X23	19/03/2019	PL	9	 Action		10:14:21	10:14:36	00:14.531
Patient X23	19/03/2019	MH	4	 Exploitation		10:13:06	10:13:21	00:15.549
Patient X23	19/03/2019	MH	5	 Exploration	i did wonder if...	10:13:21	10:13:37	00:16.399
Patient X23	19/03/2019	MH	6	Other communication	only three or four that come up. Corrects his mental model	10:13:37	10:13:50	00:13.554
Patient M4	20/11/2018	MH	1	Other communication	explains how to do specific tests.	13:05:22	13:05:36	00:14.696
Patient M4	20/11/2018	MH	2	Other communication		13:05:36	13:05:51	00:14.281

*Note.* This example is based on live coding, the start and end times of each event are displayed following 24-hour time notation.

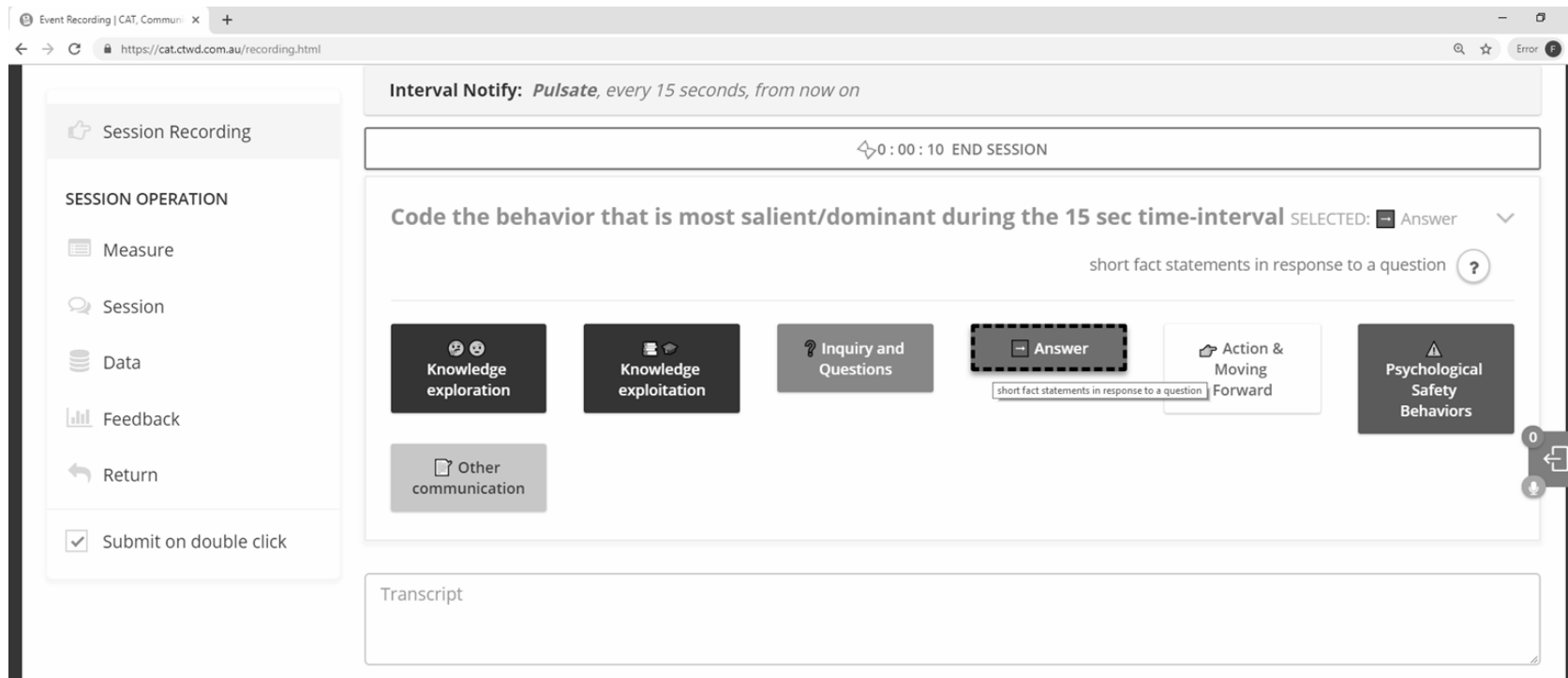


Figure 1. Screenshot and example of the coding scheme used in the exemplary research project.

Note. Navigation on the left shows different features of CAT (e.g., *Measure* links to different measures, *Session* links to data organized by different sessions). The header shows the sampling interval of 15 sec, when using the tool, the codes of the coding scheme will be displayed in different colors. Codes can be logged by double clicking.

We classified ICCs according to Cicchetti's (1994) proposed cutoff criteria: below .40 = **poor**; .40-.59 = **fair**; .60-.74 = **good**; and .75-1.00 = **excellent**.  
 Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. <http://dx.doi.org/10.1037//1040-3590.6.4.284>.

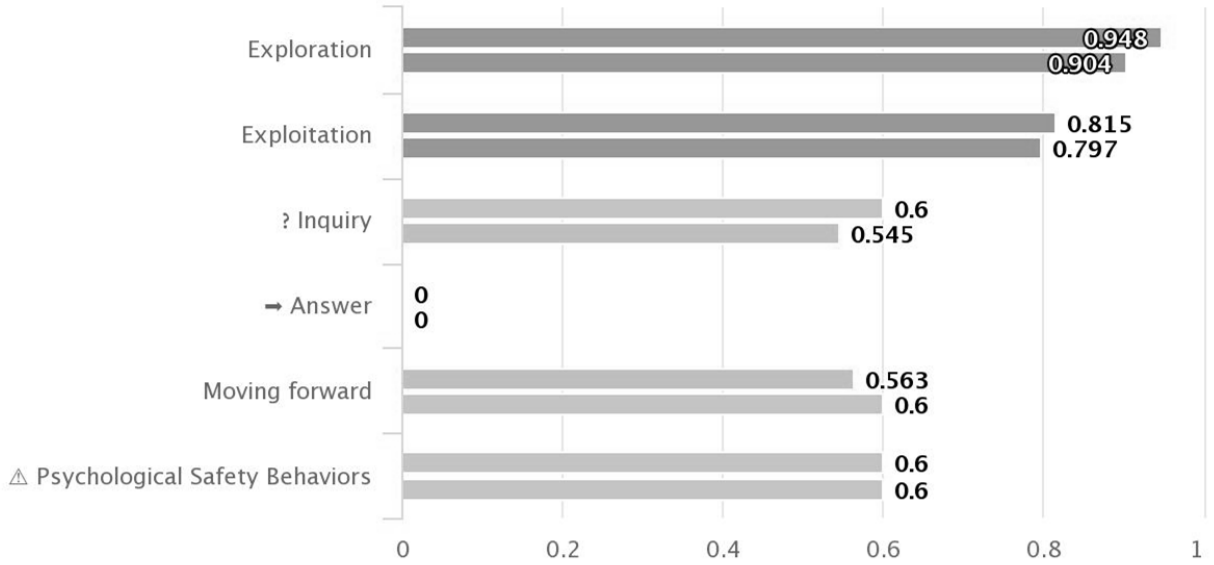
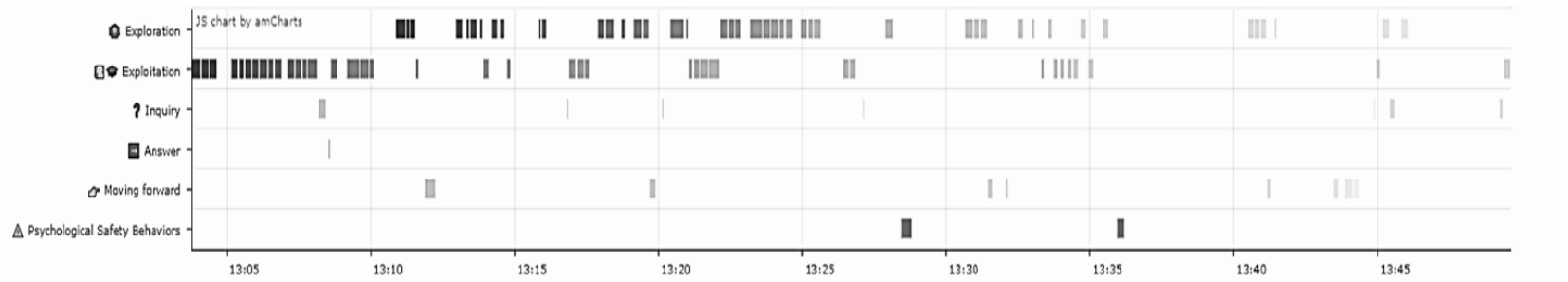


Figure 2. Visualization of interrater reliability feedback.

Note. We selected a 10 minute time window to obtain frequencies for each single code. Top line: Relative ICC (i.e., estimation of consistency in rank-order of values between both coders), Bottom Line: Absolute ICC (i.e., estimation of consistency in absolute frequency values between both coders).



*Figure 3.* Visualization of the temporal sequences of coded behaviors for one expert panel meeting.

*Note.* The course of time is shown on the *x*-axis following the 24-hour time notation (i.e., the meeting took about 50 min., starting after 13.00 and ending around 13.50), the *y*-axis displays one line for each code of the coding measure.

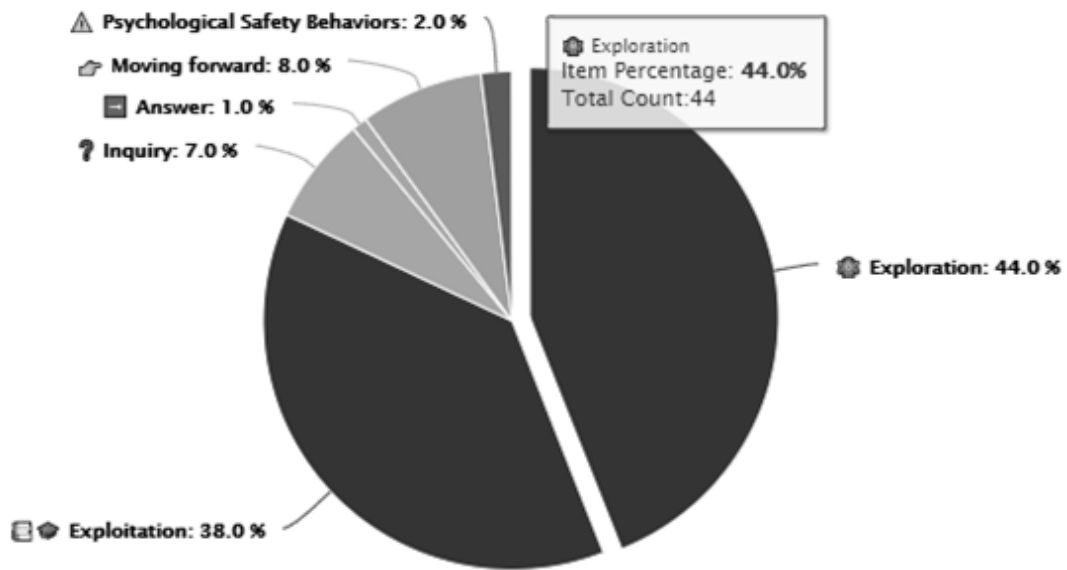


Figure 4. Team communication summary from a live observation of one expert panel meeting.



### **Author biographies**

**Florian Klonek** is a Postdoctoral Fellow at Curtin University, Future of Work Institute, Centre for Transformative Work Design, Australia. He completed his PhD at the Department of Industrial/Organizational and Social Psychology at TU Braunschweig, Germany. His research interests are in work design, team dynamics at work, change management, and leadership.

**Annika L. Meinecke** is a Postdoctoral Fellow at the Department of Industrial/Organizational Psychology at the University of Hamburg. She holds a PhD from TU Braunschweig, Germany. Her research on leader-follower dynamics, team processes, and interaction analysis has been published in *The Leadership Quarterly* and *Journal of Applied Psychology*, among others.

**Georgia Hay** is a PhD candidate at the University of Western Australia; and an associate at the Curtin University Future of Work Institute, within the Centre for Transformative Work Design, Australia. Her research interests span the topics of work design, occupational context, social identity, change management, and entrepreneurial decision-making.

**Sharon Parker** is an Australian Research Council Laureate Fellow, a Professor of Organisational Behavior at Curtin University, and the Director of the Centre for Transformative Work Design, Australia. Her research focuses on job and work design, and she is also interested in proactive behavior, change, well-being, development, and job performance.