Dissertation zur Erlangung der Doktorwürde an der
Philosophischen Fakultät der Universität Freiburg (Schweiz)

Processes and Performance in Healthcare Teams

Team Processes, Tasks and Clinical Performance in Interdisciplinary Action Teams

Jan Schmutz, Vechigen BE

Genehmigt von der Philosophischen Fakultät auf Antrag der Professoren
Tanja Manser (1. Gutachter) und Tom Reader (2. Gutachter).

Freiburg, den 17. November 2014 (Datum der Thesenverteidigung).

Dekan Prof. Dr. Marc-Henry Soulet

“It is not a question of how well each process works, the question is how well they all work together.”

- Lloyd Dobens

Danke – Merci - Graçias - Thank you to…

…Tanja; für die lehrreichen, spannenden und auch immer wieder amüsanten letzten drei Jahre, die jederzeit vorhandene Unterstützung in fachlicher sowie in sozialer Hinsicht, das immer schnelle Feedback und die guten Ratschläge

...Alexandra; für die vielen Stunden, welche sie in den letzten drei Jahren stets interessiert meinen (für einen Laien manchmal etwas weltfremden) Ideen und Erlebnissen zugehört hat und mich immer in meinem Tun bestärkt und unterstütz hat. Darüber hinaus wusste sie immer rat, wenn es darum ging meine „spannenden" Resultate in ästhetischer und verständlicher Form darzustellen.

...Johanna; für die tolle Unterstützung in meinem Projekt vor allem in der Endphase meiner Dissertation, das extrem effektive und seriöse Aufbereiten von zahlreichen Datensätzen und Excel-Dateien, das Codieren von Videos und für die immer unterhaltsame Gesellschaft im vierten Stock

...Steven und Surabhi; für ihre Unterstützung bei der Videocodierung

...Mariel; für ihre Unterstützung beim Durchschauen von zahlreichen Literatur-listen

...Laurenz; für sein Coaching in allen statistischen und methodischen Ange-legenheiten

...Sven; für seine fachlichen wertvollen Anregungen zur Dissertation und sein Verständnis für meine Nachlässigkeit in haushaltlichen Angelegenheiten in der Endphase dieser Dissertation

...Denise; für ihre professionellen sprachlichen und fachlichen Anmerkungen zur ganzen Dissertation

...allen Koautoren; für die Zusammenarbeit und Unterstützung in allen medizinischen/praktischen Angelegenheiten

Table of Contents

1. Introduction and Overview

Teams are everywhere today in the world of labour, from aviation to medicine to fire service, in management and in government, and in all kinds of professional sports. A world without teamwork is unimaginable, and teams carry out much of the work of modern life (Brannick & Prince, 1997). Even in research itself evidence shows that in the last 60 years teams of authors increasingly dominate solo authors in producing knowledge of all kinds in terms of frequency and quality (Wuchty, Jones, & Uzzi, 2007). The field of industrial and organizational psychology aims to understand the individual, his behaviour, and his perceptions in the work environment (Landy & Conte, 2009). Since individuals must work in teams in nearly any organization, it is not surprising that team research has become integral to industrial and organizational psychology.

Adverse events like airplane crashes, deaths by friendly fire due to errors in military command and control, and false diagnoses in hospitals have often been attributed to poor teamwork and thus have contributed to widespread interest in researching teams (Ilgen, 1999). But the goal of this research has not been solely to investigate effects of teamwork on individuals' wellbeing or the productivity of the team in an organization. Teams and the processes within a team are studied to prevent errors and thus to design safer environments in fields such as aviation and hospital care, among many others.

The last decade has seen an increase in literature about teams in the context of safety management in high-risk organizations. This research has centred on teams in a wide variety of industries, including aviation (Grote, Kolbe, Zala-Mezo, Bienefeld-Seall, & Kunzle, 2010), the nuclear power industry (Waller, Gupta, & Giambatista, 2004), and various healthcare sectors like surgery (Pronovost & Freischlag, 2010), anaesthesia (Burtscher, Manser, et al., 2011) and resuscitation (Fernandez Castelao, Russo, Riethmüller, & Boos, 2013), among many others.

This Ph.D. thesis contributes to this line of research by focusing explicitly on healthcare teams. This work aims to continue the existing research in this field and to contribute to a deeper understanding of team processes and the related variables (i.e. team inputs, team outputs, and the task itself).

The first part of this thesis introduces team research in industrial and organizational psychology (Chapter 2), including relevant definitions (Chapter 2.1) and models (Chapter 2.2), the role of the task (Chapter 2.3), and performance as a team output (Chapter 2.4). Chapter 3 examines the present study including an overview (Chapter 3.1) of the four scientific studies (Studies A, B, C, and D) that form the main part of this Ph.D. thesis. Chapter 3.2 links and integrates them under an Input-Process-Output (IPO) view. Chapter 4 provides a general discussion, and Chapter 5 presents the four original study articles.

Figure 1. Basic Input-Process-Output model of group processes

All four studies can be integrated into an IPO model of teams (McGrath, 1964) and will cover different aspects of this model (Figure 1). The IPO model postulates that inputs (e.g. team or task characteristics) are transformed into outputs (e.g. finished projects, healthy patients) through team processes (e.g. coordination, communication).

Study A is a systematic review of the literature and investigates the statistical relationships between processes and outputs. Study B deals with the output itself, in particular the measurement of clinical performance. Study C uses team task analysis to examine a task as well as the team process requirements a task places on the team in three different medical scenarios. Finally Study D investigates the relationship of team processes and outcomes with task type as a moderator of this relationship.

2. The Study of Small Groups in Industrial and Organizational Psychology

Originally the study of groups was mostly the province of the field of social psychology, where researchers investigated phenomena like deindividuation in groups (Diener, Fraser, Beaman, & Kelem, 1976), conformity to majority opinion within a group (Asch, 1955), or the effect of groupthink (Janis, 1972). When Levin and Moreland (1990) later reviewed small group research, they concluded that "teams are alive, but living elsewhere" (p. 620). What they meant was that research about small groups had begun to decline in the field of social psychology but was increasing in journals devoted to industrial and organizational psychology.

An early example of the study of groups within this latter field was the Hawthorne studies (Roethlisberger & Dickson, 1939), which first drew attention to the role of work groups in an organizational context and to their potential to influence productivity and attitudes of individual workers. Initially the Hawthorne studies were simple, but they grew more complex as the research began to target unanticipated social influences on the work groups and the individual (Guzzo & Shea, 1992). Social psychology usually considered groups independently from their context. In contrast, industrial and organizational psychology uses a more holistic approach that focuses on individuals, teams, and context and argues that groups always interact with their environments and that the context of a group cannot be ignored. Especially in work groups, it cannot be assumed that the same processes are present in different work environments or types of organizations. A project team developing new software, for example, is confronted with different challenges than a team of paramedics resuscitating a patient.

Guzzo and Shea (1992) use the metaphor of a cloud to illustrate the dynamics within a group as well as how it interacts with its environment. This idea is based on a discussion about social systems by Karl Popper (1972), who distinguishes between systems that function like clockwork and systems that function like clouds. Unlike clockwork, which is regular and rigid, clouds change form and size and are strongly affected by environmental conditions. It is almost impossible to predict precisely the movements of the molecules in a cloud. Guzzo and Shea (1992) argue that work groups are similar to clouds: although they have some regularity, like clouds they are disorderly and highly responsive to environmental influences. Thus, knowledge of one individual does not give us an understanding of the entire group, and the behaviour of one group member only tenuously predicts the behaviour of others. Furthermore, the behaviour of a group is not regular and predictable like clockwork; rather, variation is the norm. This variation is mostly caused by the work environment of groups (Guzzo & Shea, 1992).

This image of work groups as clouds illustrates the complexity and dynamic nature of the group and makes clear that group research in industrial and organizational psychology is challenging. This complexity results not only from dynamics within the team (e.g. group members' interaction and attitudes) but also from contextual dynamics (e.g. the adaptation of the group to embedding contexts, according to Arrow, McGrath, & Berdahl, 2000). To master these challenges and understand group performance in organizations, clear definitions of the constructs (in this case, *groups*, *small groups*, and *teams*) as well as well-founded theoretical models of teams are necessary.

2.1 Definitions

The use of the terms *group*, *small group*, and *teams* is sometimes vague and overlapping (McGrath, 1984). This section provides important definitions of the terms used throughout the thesis.

Group and small group. In the literature the use of the word *group* is sometimes vague, and different definitions exist. In industrial and organizational psychology the terms *group* and *small group* are often used as synonyms (Arrow et al., 2000; McGrath, 1984). Some confusion arises when the terms are used synonymously, since a group can be seen as a broader term than a small group (as the word "small" already implies). Groups but not small groups can include large sets of people who belong to a social category (e.g. lower-class citizens) or people in physical proximity to each other or with a common destiny who do not necessarily interact with each other (e.g. students in the same classroom, people riding a bus). Since these large sets of persons are usually the subjects of social psychology or sociology and not the subject of industrial and organizational psychology, the term *small group* might be more accurate in an industrial and organizational psychology context (Arrow et al., 2000).

Shaw (1971) defines small groups as "two or more persons who are inter-acting with one another in such a manner that each person influences and is influenced by each other" (p. 8). Hackman (1987) provides a definition of a small group, defining it as an entity that is perceived by its members and by non-members as a group; furthermore, its members have interdependencies but also a differentiation of roles and duties within the group. Arrow et al. (2000), based on the work of McGrath (1984), provide a more comprehensive definition. They define a small group as a system with loosely coupled individuals mutually interacting, interdependent members, projects, and technology with a shared collective identity. Small groups have psychological and temporal boundaries, and the members of the group are aware of the group as an entity and of their membership in it. Furthermore, members' behaviour is linked with shared consequences for the whole group (Arrow et al., 2000).

Arrow et al. (2000) as well as other researchers (Ilgen, Hollenbeck, Johnson, & Jundt, 2005) treat small groups in the work setting as complex, dynamic, and adaptive systems. Complexity theorists describe system complexity as a system with many interconnected parts and a complicated structure (Gell–Mann, 1995). As mentioned above, small groups are interdependent and have a differen-tiation of roles and duties. The more distinct the group members, their roles, and their actions are, the more complex the group is. A typical example for a complex system is a project team consisting of members of different organizations coming together for one specific project. Members have different educational backgrounds and roles (e.g. project manager, IT specialist, personnel manager), and their actions are highly linked with each other (e.g. the personnel manager telling the IT specialist what functions a specific software program needs). Small groups are considered dynamic, because they do not remain the same over time. Processes within the group change due to members' permanent adaptation to the embedded environment (Burke, Stagl, Salas, Pierce, & Kendall, 2006) or just because of automation of processes over time. Furthermore, the composition of a small group is dynamic because members who have a specific role or roles can be replaced by one or more other members who can play the same role or roles and who join the group once project implementation is underway (McGrath, Arrow, & Berdahl, 2000).

Teams. Authors use the term *team* for a group in a work context (e.g. project teams, surgical teams, police teams) whereas small groups include also teams in a non-working context. Salas, Rosen, and King (2007) define teams as identifiable work units consisting of two or more people with several character-istics including dynamic social interactions, with meaningful interdependencies, valued and shared goals, a particular lifespan, distributed expertise, and clearly assigned roles and responsibilities. One important attribute of this definition of teams is members' distinct functions, whereas in a group, members could be interchangeable and all have the same function, such as a group whose task is to solve anagrams (Brannick & Prince, 1997). However, some teams may divvy up tasks and replace specific positions. For example, the person in an aircrew who talks to air traffic control may change during the flight; however, only one person at a time is responsible for the radio.

One specific type of team is the interdisciplinary action team (IAT). In an IAT, members with different backgrounds, skills, and roles have to coordinate their actions in intense and unpredictable situations (Edmondson, 2003; Sundstrom, De Meuse, & Futrell, 1990). By definition IATs must respond to unexpected incidents in a coordinated way, often requiring an open transfer of information to make real-time coordination of actions possible. IATs typically work in high-risk environments like aviation, healthcare, or the power industry. They are formed ad hoc and work together for a specific task (e.g. a surgical operation), and then the team dissolves once a job has been completed. The present thesis investigates typical IATs in different healthcare settings like cardiac anaesthesia, live trauma resuscitation (as shown in Study C), and paediatric emergency (as shown in Studies B, C, and D).

2.2 Theoretical models for understanding teams based on an IPO approach

This chapter aims to present theoretical models for understanding teams and the processes occurring within a team.[1] There have been a variety of schools of thought on teams and groups.[2] This chapter focuses on theories based on the IPO (Input-Process-Output) model that has dominated team research and theory to the present day (McGrath, 1964). The IPO model was also the underlying framework for all four studies in this thesis. Furthermore, this chapter explains two specific variables that are part of the IPO model in more detail, because they play an important role in the following studies: clinical performance as an output (Study B) and the role of the task (Study C and D).

Input typically refers to the things team members bring to the group, including status, personality attributes, level of experience, demographic attrib-utes, and others. These inputs are often referred to as team characteristics. Other inputs are the characteristics of the task, like the type of task (e.g. routine vs. non-routine) or the resources a specific task brings with it (Guzzo & Shea, 1992). The *process* refers to the interactions among team members that transform inputs into outputs (e.g. exchange of information, distribution of tasks). *Output* then refers to the product resulting from the team's work. This product can include ideas, decisions, a finished project, or the successful treatment of a patient. The IPO model in this basic formulation is explicitly causal and rather static. Output is a consequence of the nature of team processes, which themselves result from inputs brought into the team setting (Guzzo & Shea, 1992).

Although this static IPO model governed much early research in the field, researchers have developed alternative models that consider team processes as more dynamic systems. Hackman (1987) formulated two alternative models. He states that processes do not mediate all input variables, although there is a direct line from inputs to outputs. For example, the work experience of one team member could lead directly to a good output without mediation by team processes. Hackman (1987) further proposes a

---

[1] The terms *model* and *framework* are used interchangeably throughout this thesis. The same terms as in the corresponding reference will be used.

[2] For a comprehensive review, see Guzzo and Shea (1992) and Arrow et al. (2000).

model that deviates even more from the basic IPO formulation whereby inputs influence both processes and outputs; in this case, team processes are not integral to explaining team outputs. But today such a model would not pass muster, since numerous studies have scientifically proved a relationship between team processes and team outputs (DeChurch & Mesmer-Magnus, 2010; Schmutz & Manser, 2013).

Kozlowski, Gully, Nason, and Smith (1999) propose the theory of compilation and performance, in which time becomes an important factor: input, process, and output develop over time as the team as a whole interacts with its environment and with the team members themselves. Knowledge, attitudes, and behaviours in a developmental stage are both inputs and processes that impact team performance (output). The whole process is not one but several reciprocal cycles where the output influences the input of a subsequent cycle. Processes are not static, but rather unfold over time.

Marks, Mathieu, and Zaccaro (2001) extend the classical IPO model and provide a more detailed framework for team processes. They define team processes as members' interdependent acts that convert input to output through cognitive, verbal, and behavioural activities directed toward organizing taskwork to achieve collective goals. *Taskwork* in this case is defined as the teams' interaction with tasks, tools, machines, and systems. Thus team processes are used to direct, monitor, and align taskwork to achieve a meaningful outcome. The central point of this conception of team processes is the interaction of team members with each other and with the task environment. The authors state that some research investigating IPO relationships has considered constructs like collective efficacy, potency, cohesion, and situational awareness as team processes. They argue, however, that these constructs do not represent actual interaction processes in the team but rather describe qualities of a team that reflect members' attitudes. Marks et al. (2001) defined these types of variables as "emergent states" that need to be distinguished from actual team processes in the IPO model. The authors' understanding of a more dynamic view of team processes aligns with Kozlowski et al. (1999).

Researchers traditionally adopted a static view of the IPO model and investigated IPO relationships within one task, without considering temporal influences. Time is always a component in processes involving work teams when striving toward a common goal (Locke & Latham, 1990). Through successfully reaching a goal, a team gains experience or develops new routine actions over time that function as inputs in future tasks. Based on these thoughts Marks et al. (2001) propose that the basic formulation of the IPO model is inadequate to represent the entire life cycle of a team project. Instead a team passes through several episodes where IPO cycles run sequentially and simultaneously and where the output serves as input for the next cycles. In addition, the authors define two types of episodes where IPO cycles run: an action phase during which teams engage in acts that contribute to accomplishing a goal and a transition phase during which teams primarily focus on evaluating and/or planning activities. Depending on the episode, a team performs different kinds of team process behaviour.

Ilgen et al. (2005) see three types of limitations in the basic IPO model. First, interactions have been documented between inputs and processes as well as between different processes. Such interactions contradict the linear view of the basic IPO model. Second, like others before them (Kozlowski et al., 1999; Marks et al., 2001), they criticise the single-cycle path of the model. Although the authors of the classical work mention the possibility of different feedback loops within the model (Hackman, 1987), it should be seen more as a multi-cycle model where outputs like team performance are treated as inputs for future team processes. Third, the term “team processes” itself limits research, because constructs often described in the literature are emergent cognitive or affective states that lack the character of a process, as Marks et al. (2001) previously criticised.

As an alternative to the classical IPO model, Ilgen et al. (2005) suggest instead the Input-Mediator-Output-Input (IMOI) model. Replacing *process* with *mediator* reflects a broader range of variables that can mediate the relationship between inputs and outputs (including emergent states). The extra “I” (for input) at the end invokes the notion of cycling causal feedback in terms of outputs that will serve as inputs for a future team process. Furthermore, the authors explicitly reject using hyphens between the letters (IMOI) as some authors have done when identifying the IPO model (as I-P-O). Their refusal to use hyphens signifies that the linkage between the variables may not be linear.

Burke et al. (2006) present another alternative, the Input-Throughput-Output model, which is based on an IPO model but focuses on the dynamic abilities of a team to adapt to its environment. The aim of the model is to explain *adaptive team performance* as the dependent variable of interest. The authors define adaptive team performance as an emergent phenomenon that develops over time as a result of recursive cycles whereby one or more team members use their resources to functionally change current cognitive or behavioural goal-directed actions or structures to meet expected or unexpected demands. The model describes individual characteristics (i.e. knowledge, attitudes, traits, and abilities) and job design characteristics as inputs. The *throughput* phase is an adaptive cycle in which the team assesses the situation, formulates a plan, executes the plan, and then learns from the process. In addition, the throughput phase includes emergent states (i.e. shared mental models, situational awareness, and psychological safety) that influence the adaptive cycle as proximal outcomes as well as inputs for the next cycle. This condition explains the replacement of the term *process* with the term *throughput* in this model, because the throughput phase also includes emergent states that do not show characteristics of a process per se, as Ilgen et al. (2005) and Marks et al. (2001) already stated before. The ultimate result of the interaction of the adaptive cycle and emergent states in the throughput phase is team adaptation. Team adaptation in turn is linked through feedback loops with inputs again, which is in line with previous theories that see the IPO model as a dynamic multi-cycle framework where outputs of the previous cycle influence the inputs in a subsequent cycle (Ilgen et al., 2005; Marks et al., 2001).

As outlined above there is a trend toward more complex theoretical models for understanding and representing teams and their processes. Although these models differ in specific aspects, they reflect the underlying notion that teams are complex, dynamic systems, existing in larger systemic contexts of tasks, people, technologies, and settings (Ilgen et al., 2005). This notion is in line with other theoretical models not based on the IPO model.

2.2.1 Other theoretical models of teams

Arrow et al. (2000) present a general theory of small groups (including teams) as complex, adaptive, dynamic systems. They present a theory based on concepts borrowed from several other fields like systems theory, dynamical systems theory, complexity and chaos theory. The authors state that throughout a group's life three levels of causal dynamics—local, global, and contextual—constantly form the group.

Local dynamics refer to the activity of a group's constituent local variables. That involves the activity of the system parts and the rules that govern these activities. These rules include implicit norms or procedures (e.g. stereotypes, intentions) or more explicit norms like the distribution of different tasks according to the team members' background. The central concept of local dynamics is coordination (i.e. coordination of actions, coordination of understanding, and coordination of goals). In the first type, actions are coordinated among team members so that all actions fit together into an intended spatial and temporal pattern. Coordination of understanding, the second type, means achieving an agreement among group members regarding more cognitive processes. That means coordinating information about the process, norms, and division of labour to achieve the goal. The third type of coordination, coordination of goals, includes the mutual adjustment of individual purposes, interests, and intentions among group members.

Global dynamics are rules and activities on a systemic level emerging out of local dynamics. The term describes the state of a group as a whole. The literature often refers to *global dynamics* as group development investigating variables like group performance, cohesiveness, or conformity.

Last, c*ontextual dynamics* are described as the features of groups that embed context that shapes and constrains global and local dynamics. A small group has to adapt to a variety of contextual influences like adding or removing new projects within a team, changing its composition, or allocating resources to different teams in an organization. A team is constantly interacting with its environment and thus has to adapt to it if necessary (e.g. a surgical team has to adapt the team organisation to the size and arrangement of the operating room).

2.3 The role of the task

The task itself is a central variable in team research and appears in the literature based on an IPO approach in either of two ways (Guzzo & Shea, 1992). First, as already stated, task characteristics can be interpreted as inputs that determine the team processes necessary to master a specific task. Second, a task can be interpreted as a moderator of the relationship between process and output. Depending on the nature of the task, specific team processes may be effective and lead to good outcomes or not. Statistical models with task as a moderator between team process variable and an output variable already have been tested (DeChurch & Mesmer-Magnus, 2010; Schmutz, Hoffmann, Heimberg, & Manser, under review).

Teams that are dynamic, adaptive systems not only have to adapt the team processes to contextual influences or cues (Arrow et al., 2000; Burke et al., 2006) but also to the task itself. A task involving resuscitating a patient demands different processes from a team than a routine check of a patient does. In this context authors talk about task adaptive behaviour (Tschan, Semmer, Nägele, & Gurtner, 2000). Thus it is important to identify the team process requirements that a specific task demands from a team (Parker, Schmutz, & Manser, under review).

Various taxonomies of team tasks exist in the literature (McGrath, 1984; Steiner, 1972). For example, Steiner (1972) proposes the following sets of distinctions. He distinguishes between unitary tasks that require no single group output and divisible tasks that do require one. Additive tasks require that the groups' resources get pooled together, whereas disjunctive tasks require that only one member does the task for the group. In conjunctive tasks all group members must individually succeed, and in discretionary tasks resources can be combined in any way to reach the common goal.

Because tasks in organizations are much more complex than Steiner's task categories imply, they have seen no use in organizational settings (Guzzo & Shea, 1992). They may be useful in a standardised laboratory setting but of little help in theoretical research in the field. These taxonomies fail to describe the work that teams in real-life organizations do. Taxonomy of tasks must be able to create categories of similar tasks but still reflect the specific characteristics of each task. Teams in different organizations have to deal with very different tasks. Due to the different nature of tasks, it might be difficult to develop a task taxonomy that can be used for project teams as well as surgical teams. Thus taxonomies should be developed that are domain-specific. This is only possible if the researcher thoroughly understands the investigated task through team task analysis (Annett, 2003; Parker et al., under review). Such an analysis enables the understanding of the task and the similarities as well as the differences among tasks so as to help develop suitable and reasonable task taxonomies.

2.4 Performance as team output

Measuring the performance of a team is important for research testing the process-output relationship as well as to evaluate teamwork training. Team training literature often focuses on the quality of specific team process behaviours like backup behaviour, communication, or team monitoring (Dickinson & McIntyre, 1997). This approach is appropriate if it is well known what process behaviours are effective in a given situation. To identify what team processes are effective in what situations, each process first has to be linked with a performance outcome. To do this, team processes (e.g. giving feedback, distributing tasks) should be assessed objectively without any evaluation and linked to objective performance measures that identify good performance in a specific team task. But what is good performance, and how can it be assessed?

Job performance is a core concept in industrial and organizational psychology (Campbell, 1990; Nerdinger, Blickle, & Schaper, 2011; Viswesvaran, 2001). The importance of this construct is reflected in the large volume of literature concerning job performance and particularly job performance assessment. Job performance data can be used for administrative purposes including salary administration, promotions, or layoffs. It can also be used as a basis for feedback by identifying individual strengths and weaknesses and also for research purposes that include evaluation of interventions or validation of a new selection technique (Cascio, 1987).

The literature defines job performance in various ways. Because there are no universal measures of performance for all jobs, performance in a given job is measured in the context and is closely it related to a company's goals (Campbell, 1990). For example, the medical staff of a hospital will behave differently if management communicates a goal, such as to reduce treatment costs or to increase patient satisfaction. Campbell (1990) defines performance as follows: “Performance is behavior. It is something that people do and is reflected in the actions that people take” (p. 704). This definition, which focuses on action, can be seen as a behaviour-centred view of performance. On the other hand authors agree that performance also has outcome-related aspects (Anderson, Ones, Sinangil, & Viswesvaran, 2001; Nerdinger et al., 2011; Sonnentag & Frese, 2002). The outcome-related aspect refers to the results of one’s behaviour. In our example, an outcome-related aspect could be the treatment costs or the percentage of satisfied patients after one year during which the medical staff members have modified their behaviour.

This distinction is important for performance assessment: one can rate the behaviour of a team and/or the outcome of this behaviour. The rating of performance behaviour can be achieved through observation by experts in the field. An expert knows the critical behaviours for high performance in his area (Nerdinger et al., 2011). Furthermore, specific rating systems or checklists that include what critical behaviour is expected to be performed can be useful in-struments to assess the behaviours related to performance of an employee or a team (Donoghue et al., 2011; DuPaul, Rapport, & Perriello, 1991).

Performance outcomes depend highly on organizational context and can be very specific. Examples for performance outcomes could be number of sales, number of publications, turnover of employees, number of successful heart operations by a surgical team, or reduced patient mortality in a hospital. The problem of outcome-related aspects is that they are hard to attribute conclusively to changes in team members’ behaviour (Campbell, 1990; Ner-dinger et al., 2011; Sonnentag & Frese, 2002). Outcome performance measures are related to the result of the actions and depend on more than just individuals’ behaviour. Of course, a relationship exists between the behaviour and the outcome, but these two metrics never overlap completely. Nerdinger et al. (2011) refer to this as the “attribution problem” and states that performance should be measured exclusively by the behaviour criteria and not by the outcome, because confounders influence the outcome.

For example, a sales consultant could accomplish a consultation with a potential customer perfectly, but if the customer has no money or gets a better offer from another company, he will never buy the product. In another example, the survival of a resuscitated patient is better predicted by factors such as the duration of the patient's cardiac arrest, the type of heart arrhythmia, and the patient's age than by the performance of the clinical team performing the resuscitation (Cooper & Cade, 1997).

Thus in these cases it would be more accurate to measure the actions and not the outcome. Of course, good process performance should be and is related with outcome performance. But in field research it is often impossible to control for all the confounding variables, and thus it might be difficult to interpret a negative outcome because it is unclear if it resulted from the incorrect behaviour of the team or from confounding variables.

2.4.1 Performance measures in healthcare

In healthcare literature many different performance measures are used. Typical examples are mortality, morbidity (Davenport, Henderson, Mosca, Khuri, & Mentzer Jr, 2007), length of hospital stay (McCulloch et al., 2009), task execution time (Catchpole et al., 2007), or ratings of specific behaviours according to medical guidelines (Burtscher, Kolbe, Wacker, & Manser, 2011). Especially in healthcare, it is hard to attribute specific outcomes to the behaviour of an individual or a team. The success of patient treatment relies on many aspects, and it is hard to control all potential factors.

Due to this fact it is common in healthcare to assess the behaviour aspect of performance with tools like checklists based on medical guidelines (Donoghue, Nishisakia, Suttona, Halesc, & Boulet, 2010), global rating instruments, time-to-event assessments, or achievement of key steps/goals (Holmboe & Hawkins, 2008). Performance based on behaviour aspects can be assessed in almost every setting. During a medical simulation, where it is hardly possible to assess patient-related outcomes, behaviour performance measures are preferable.[3]

[3] In medical simulations an actor or a mannequin that does not show actual symptoms replaces the patient.

3. The Present Study

The aim of this thesis is to build on existing research on teams in healthcare and to contribute to a deeper understanding of team processes and the related variables, especially clinical performance and the team task. The four studies included in this thesis investigate various aspects of the IPO model of teams. This model has already been the basis for studies in healthcare. Based on a IPO model Reader, Flin, Mearns, and Cuthbertson (2009) provide the team performance framework for the intensive care unit (ICU; Figure 2). The authors specify inputs as team characteristics (e.g. team climate, team hierarchies), task characteristics (e.g. time pressure, protocols for completion of tasks), and leader characteristics (e.g. leadership style, personality). They specify four team processes as team communication, team leadership, team coordination, and team decision-making that lead to two types of outputs: patient outcomes and team outcomes. Team outcomes include all outcomes concerning the team or team members as a result of the team process (e.g. stress, job satisfaction). Patient outcomes include the results of the team processes concerning the patient (the result of the treatment by the team). As in other IPO theories presented earlier, a feedback loop links the output with inputs, implying a multi-cycle understanding of the framework.

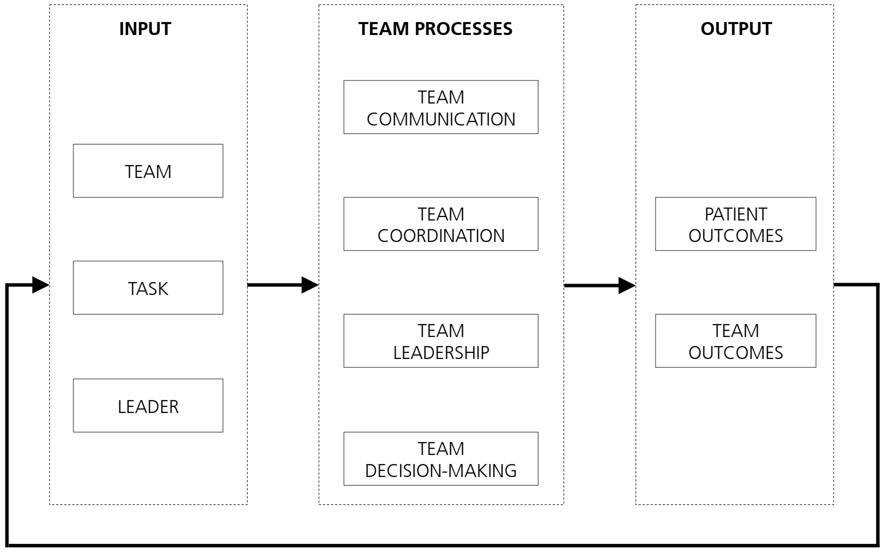

Figure 2. Team performance framework (Reader et al. 2009).

This framework, although developed for an ICU environment, can be used for other healthcare environments where teams are treating a patient (e.g. surgery, paediatrics). Thus this IPO framework serves as the basic theoretical model for this thesis, which investigates multiple variables and relationships concerning it.

3.1 Summary of thesis studies

3.1.1 Study A

Do team processes really have an effect on clinical Performance? A systematic literature review

Jan Schmutz and Tanja Manser (2013)

The first study is a systematic literature review summarizing articles that examined the impact of team process behaviour on clinical performance in healthcare teams. There has been other literature reviews indicating that team processes play an important role in medical teams (Manser, 2009; Reader et al., 2009). To date this is the only literature review that focuses solely on statistical relationships between the two variables in a variety of healthcare teams.

A literature search in five major databases was conducted. Inclusion criteria were: Peer-reviewed papers in English published between January 2001 and March 2012, which showed or tried to show a) a statistical relationship of a team process variable and clinical performance or b) an improvement of a performance variable through a team process intervention. Study quality was assessed using predefined quality indicators. For every study we calculated the relevant effect sizes. Team processes included in the study were defined according to Reader et al.'s framework (Reader et al., 2009).

Twenty-six studies were included in the review, six of which were intervention studies. Every study reported at least one significant relationship between team processes or between an intervention and performance. Some nonsignificant effects were reported. Most of the reported effect sizes were medium or large. The study quality ranged from medium to high. The studies were highly diverse regarding both the specific team process behaviours investigated and the methods used. However, they suggest that team process behaviours do influence clinical performance and that training results in increased performance. Future research should rely on valid and reliable methods to assess processes such as teamwork or coordination and focus on developing adequate tools to assess process performance in the clinical setting, linking them with high-quality outcomes.

3.1.2 Study B

Five steps to develop checklists for evaluating clinical performance: An integrative approach

Jan Schmutz, Walter J. Eppich, Florian Hoffmann, Ellen Heimberg, and Tanja Manser (2014)

Study A outlines the high diversity in team processes and investigated outcomes. Based on the premise that valid team process measures are as important as valid outcome measures. Study B focuses on clinical performance assessment and more particularly on the process of developing checklists for evaluating clinical scenarios. The process of developing checklists to rate clinical performance is essential to ensure their quality, thus an integrative approach for designing checklists that evaluate clinical performance has been developed.

The approach consists of five predefined steps. Step 1: Based on the relevant literature and their clinical experience, the authors drafted a preliminary checklist. Step 2: The authors sent the draft checklist to five experts who reviewed it using an adapted Delphi technique (Gordon, 1994). Step 3: Three scoring categories were devised for items following pilot testing. Step 4: To ensure that the changes made after pilot testing were valid, the checklist was submitted to an additional round of Delphi reviews. Step 5: To weight items needed for accurate performance assessment, ten paediatricians rated all checklist items in terms of their importance on a scale from 1 (not important) to 5 (essential).

Study B used the example of a checklist for a simulation scenario of septic shock in an infant. Inter-rater reliability and validity were calculated. The five-step approach resulted in a valid, reliable tool and proved to be an effective method to design evaluation checklists.

This approach integrates published evidence and the knowledge of domain experts. A robust development process is a necessary prerequisite of valid performance checklists. Establishing a widely recognised standard for developing evaluation checklists will likely support the design of appropriate measurement tools. This approach provides a method to design a tool to rate the clinical performance of a team. This approach further formed the basis for a valid and reliable team performance assessment in Study D.

3.1.3 Study C

Coordination in healthcare action teams: Utilizing expert understanding of task and team performance requirements

Sarah Henrickson Parker, Jan Schmutz, and Tanja Manser (under review)

Study C focuses on team coordination as a team process and on the task itself, where coordination is needed. Expert action teams in healthcare are often confronted with situations that are dynamic, require real-time reactions, and must be continually assessed and reassessed. These situations require coordination among team members to achieve task goals. Thus the goal of this study is to identify specific coordination requirements for three clinical contexts: 1) clinical work in cardiac anaesthesia, 2) medical emergency with a clear treatment protocol in paediatrics (i.e. sepsis scenario), and 3) live trauma resuscitations.

Each part of Study C used task analysis to define the team tasks and to then elicit expert knowledge on coordination requirements. Comparing and collating the findings across these diverse clinical settings, the study revealed that expert teams must 1) continually reassess their current state and goals using multiple, sometimes overlapping work templates, 2) coordination requirements must be made explicit at certain points in a clinical task, though not every clinical task requires "anchors," and 3) the occurrence of a non-routine situation requires explicit coordination.

To date, research on healthcare action teams has focused on detailing the exact nature of coordination. This study takes the next step in understanding coordination by synthesising the results of coordination studies in three different healthcare action team environments. Defining critical coordination requirements without an understanding of the exact nature of the task is of limited utility. Rather, coordination requirements are highly situation-dependent in action teams. Therefore, further refined analytic approaches that combine temporal variability and task variability with coordination assessment are necessary.

3.1.4 Study D

Effective Team Coordination in Emergency Care: The Moderating Role of Task Type

Jan Schmutz, Florian Hoffmann, Ellen Heimberg, and Tanja Manser (2014)

Study D, the fourth study, focuses on the relationship between team process behaviour and clinical performance, with task type as a moderator. Like Study C, this article focuses on team coordination as a predictor for clinical team performance. Study A has shown that authors investigating coordination report different and sometimes even contradictory results. One reason for these diverse results might be the task in which coordination is investigated. Study C already stated that coordination is task dependent, and different tasks place different coordination demands on the team. Studies by other researchers have also suggested that teams have to adapt the coordination process to the task (Tschan, Semmer, Hunziker, & Marsch, 2011), but until now no consensus exists about what coordination behaviours are effective in what tasks.

Three coordination behaviours (i.e. task distribution, acknowledgement, and provide information without request [PIWR]) were investigated in 68 interprofessional teams of fully qualified clinicians during simulation-based in-situ paediatric emergency training. Based on task analyses the four training scenarios were grouped into rule-based (i.e. focused on task execution and mostly based on predefined algorithms) and knowledge-based tasks (i.e. focused on gathering and integrating information to establish diagnosis) according to Rasmussen's model for human performance (Rasmussen, 1983). Hierarchical regression analysis was used to investigate the relationship of task distribution, acknowledgement, and PIWR to clinical performance, with task type as a moderator.

Task distribution is related with clinical performance ($\beta = .24$, $p = .04$) but no moderation of task type ($\beta = -.06$, $p = .74$), and acknowledgment is moderated by task type ($\beta = -.49$, $p = .02$). Additional simple slope analysis revealed that acknowledgement is related with clinical performance only in rule-based tasks ($B =$

2.09, *SD* = .59, $p < .01$). PIWR was not related with clinical performance ($\beta = -.08$, $p = .41$), nor was this relationship moderated by task type ($\beta = -.10$, $p = .59$).

Explicitly distributing tasks seems to be important in different kinds of emergency tasks, because it ensures that all treatment actions are performed. Acknowledgement closes the communication loop, giving feedback to the sender that information or orders have been understood. This is especially important in rule-based tasks, in which the focus is on the distributed tasks that can be acknowledged, thus resulting in more effective coordination. These results are in line with similar concepts commonly taught during medical communication training (Härgestam, Lindkvist, Brulin, Jacobsson, & Hultin, 2013). PIWR should be investigated with an emphasis on the broader context of the task and its goals. We propose that PIWR can be effective under some circumstances but also can be distracting, depending on the personnel makeup of the team. Overall the results suggest that task distribution and acknowledgement are important coordination behaviours that are related with clinical performance and thus may contribute to patient safety. Furthermore, this study has important practical implications for training teams in healthcare.

3.2 Integrating and linking studies under an IPO view

All four studies cover different aspects of the IPO model for teams. Studies A and C focus more on the P-O relationship, Study B on the output itself, and Studies C and D on processes related to the task. The following paragraph states the contribution of each study to a better understanding of healthcare teams under an IPO approach. Figure 3 provides an illustration of all four studies integrated in the IPO model according to their research topic.

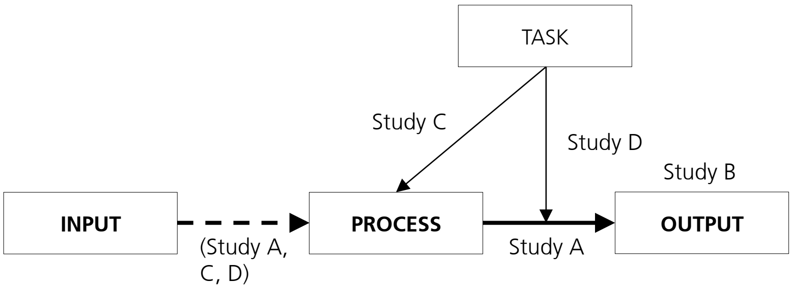

Figure 3. Studies integrated in the IPO model[4]

3.2.1 Team inputs

Team inputs are not the main focus of this research (dashed arrow in Figure 3). Nevertheless team processes cannot be investigated independently from inputs. Team processes always happen in a certain environment (with specific task characteristics) and employing a certain team (with specific team characteristics) that may potentially influence the team process. The nature of the process is grounded and bounded to specific inputs that are critical in the organizational setting (Ilgen, 1999). The specific setting in which a study has been carried out serves as an important input variable that should not be ignored.

Study A focuses on statistical relationships between a team process variable and an outcome variable. Input factors like characteristics of the investigated sample (team size, education) as well as the environment in which the study has been conducted (e.g. the ICU, surgery) are listed and included in the interpretation of the results. Study C investigates team coordination requirements in three different healthcare contexts, concluding that different tasks and their different characteristics imply different coordination requirements. Study D investigates the task as a moderator and not as an input. Nevertheless team characteristics (e.g. team size, leader experience) and the duration of a task (as a task characteristic) were assumed to influence team process behaviour and clinical performance. For this reason these variables were included as control variables in the regression model.

[4] Please note: Figure 3 is an illustration and should not be seen as a methodological model as a whole

3.2.2 Team processes

Team processes have a central role in this thesis. Study A first provides an overview about the literature investigating team processes, as defined in the team performance framework (Reader et al., 2009), in relation with outcomes in healthcare teams. Studies included in the review investigated the following process constructs: team coordination, adaptive coordination, team monitoring, leadership, communication, nontechnical skills, explicit reasoning, and teamwork. Studies C and D both focus on team coordination, Study C in a more general way in relation to the task whereas Study D investigated in particular the three coordination behaviours: feedback, acknowledgement, and task distribution.

3.2.3 Team outputs

All four studies in this thesis focus on the safety of the patient and thus include only patient outcomes and no team outcomes. Study A provides an overview of the different patient outcome measures used in the team process literature and includes process performance as well as outcome performance. Studies B and D then focus solely on process performance that is also interpreted as a patient outcome. By definition this interpretation might be incorrect, but it can be assumed that a good process performance measure correlates strongly with patient outcomes.

The development of a good performance measurement tool (performance checklists) is the main focus of Study B. Performance checklists take into account the important actions for a specific treatment and evaluate those actions across the whole process, providing a valid and reliable method to assess performance if developed systematically (Stufflebeam, 2000). This includes a theoretical foundation and an integration of official guidelines and the experiences of several experts. Based on many years of experience, experts tend to know which behaviour is important for the success of patient treatment. All these factors are integrated in the development approach presented in Study B. The performance measures in Study D resulted in a checklist rating with lists developed according to the approach developed in Study B. No patient outcomes were assessed, because Study D was conducted in a setting with a simulated patient, thus process performance assessment was the most appropriate way to measure performance.

3.3 Author's contribution to thesis articles

Table 1 lists Ph.D. candidate's contribution for *study A-D*.

TABLE 1
Overview of author's contributions to thesis articles

Paper	Ph.D. candidate's contribution
Study A Schmutz & Manser (2013)	• Substantial contribution to conception and design • Substantial contribution of acquisition of data • Substantial contribution to data analysis and interpretation • Substantial contribution to drafting and revising the article for important intellectual content
Study B Schmutz, Eppich, Hoffmann, Heimberg & Manser (2014)	• Substantial contribution to conception and design • Substantial contribution of acquisition of data • Substantial contribution to data analysis and interpretation • Substantial contribution to drafting and revising the article for important intellectual content
Study C Parker, Schmutz & Manser (under revision)	• Substantial contribution to conception and design • Substantial contribution of acquisition of data (Sub-study in paediatric emergency) • Substantial contribution to data analysis and interpretation (Sub-study in paediatric emergency) • Supportive contribution to drafting and revising the article for important intellectual content
Study D Schmutz,	• Substantial contribution to conception and design

Hoffmann, Heimberg & Manser (under revision)	• Substantial contribution of acquisition of data • Substantial contribution to data analysis and interpretation • Substantial contribution to drafting and revising the article for important intellectual content

4. General Discussion

This thesis builds upon existing research on healthcare teams and aims to contribute to a deeper understanding of team processes and relevant variables, in particular the task as well as the clinical performance as an output. The IPO model provided a useful framework for this thesis, allowing the research and its results to be integrated.

Study A revealed, among other things, that most studies investigating team processes lack a theoretical model or a concept common to all researchers, which makes comparing the results among studies difficult. A group of unrelated studies may investigate variables under the same construct (e.g. communication), but the researchers define the variables differently from study to study. For example, Davenport, Henderson, Mosca, Khuri, and Mentzer Jr (2007) defined teamwork as a climate variable characterized by a *perceived* quality of collaboration among team members, whereas Siassakos et al. (2011) assess teamwork within a simulation as a *concrete* behaviour that team members exhibit. These variations strongly suggest a need within the field of industrial and organizational psychology for clear definitions and consistent use of terms and theoretical concepts.

The four studies in this thesis incorporate a theoretical approach based on the IPO model. Such a model has proved essential to study design and planning as well as to comparing and interpreting results. Although the need to ground research in the field of industrial and organizational psychology in theoretical models may seem self-evident, we cannot say the same for a large portion of articles in the field of medicine, which often take a highly practical approach. But theory and practice should not be mutually exclusive, and a compromise needs to be found to advance research in this field.

Study A of this thesis investigates team processes in general. The construct of team processes (defined as all interactions going on in the team) is a broad term. The team performance framework provides more concrete ideas about what team processes include (i.e. coordination, leadership, communication, decision making). However, these four concepts, as formulated by Reader, Flin, Mearns, & Cuthbertson (2009), are not exclusive. Giving orders to one's team members can be defined as "leadership," but can also be interpreted as a way to communicate with the team members or to coordinate their actions. Marks, Mathieu, and Zaccaro (2001) provide a more detailed taxonomy of team processes, including 10 team process dimensions divided into transition processes (i.e. planning, goal specification, strategy formulation), action processes (i.e. progress, action and team monitoring, coordination), and interpersonal processes (i.e. conflict and action management, confidence building).

Whereas Study A investigates team processes in general, Studies C and D focus more specifically on team coordination. In the taxonomy of Marks et al. (2001), coordination is understood as an action process that orchestrates the sequence and timing of actions only. More recent studies, as well as this thesis, that investigate coordination have adapted a broader definition of coordination (Burtscher et al., 2011; Manser, Howard, & Gaba, 2008). In this broader definition, coordination is not limited to actions but also includes marshalling information. Coordination of information includes processes from other dimensions of the taxonomy of Marks et al. (2001), such as planning and strategy formulation. This broader definition is also in line with the interpretation by Arrow, McGrath, and Berdahl (2000), who argue that coordination is the central dynamic within a team. The authors identify three types of coordination into which all types of coordination within a team process would fall: coordination of actions, coordination of understanding, and coordination of goals. Coordination of information, for example, would fall under the category of coordination of understanding.

The four studies in this thesis have demonstrated that the task plays an important role in team processes. In particular, Study C demonstrates that the task not only elicits specific coordination requirements as an input, but can also act as a moderator, defining what process behaviour is effective in what kind of task. As noted earlier, industrial and organizational psychology aims to understand the individual (or a team), his or her or its behaviour and cognitions in the work environment. Of course, the task is central to the work environment, and thus researchers need a profound understanding of the specifics of whatever task they use to investigate team processes. This thesis has shown that analysing the task a team must undertake is essential to assessing and interpreting team processes in a given setting. Analysis of the task itself not only helps us to understand the nature of the task, but also provides essential information about actions, coordination requirements, cues, and work processes within the team.

4.1 Limitations and outlook

Although theories centred on the IPO model propose a more dynamic and multi-cycle approach, most studies (including this thesis) about team processes in healthcare employ a rather static interpretation of the model. Usually researchers investigate team processes within a single cycle of task per-formance, and thus the feedback loop of the output to the input of a future IPO cycle is ignored, even studies gathering team process data over a long period aggregate data into a summary index that displays only one process-outcome relationship (Marks et al., 2001). One step toward interpreting team processes more dynamically lies in research about adaptive team behaviour in which processes change depending on the stage of a task or, alternatively, in which different processes are effective at different stages of a task (Burtscher, Wacker, Grote, & Manser, 2010; Grote, Kolbe, Zala-Mezo, Bienefeld-Seall, & Kunzle, 2010).

Furthermore, future studies should investigate team processes and their development over the long term. Riethmüller, Fernandez Castelao, Eberhardt, Timmermann, and Boos (2012) published one of the first studies investigating clinical teams over a long period (specifically, on the development of adaptive coordination mechanisms). They found that in routine phases of a treatment the amount of explicit coordination over time decreased, whereas implicit coordination increased due to the development of routines within the team.

The study of Riethmüller, et al. (2010) was conducted using six teams of medical students. The authors investigated the development of coordination mechanisms within the same teams during four different training sessions. Over this time the processes within each team evolved as team members accumulated task and interactive experience. A criticism of this study could be that in real medical teams, the team members often change with every new task. A surgical team almost never has an identical constellation of members every time; sometimes the members of different sub teams (e.g. nurses, surgeons) barely know each other. We can therefore conclude that medical teams with changing constellations of members do not develop specific processes for their team in the way that an experienced sports team does (e.g. team-specific strategies or moves). That said, through repeated practice every team member gains more experience using technical skills (task experience) as well as nontechnical, team-process-centred skills (interactive experience).

Experience can influence team processes in two ways. First, team members can directly learn effective team processes over time. For example, a resident can learn to use direct and clear communication or can acknowledge orders to prevent communication errors. Second, team members develop better technical skills over time and thus free up cognitive resources to engage in team process behaviour. Team monitoring and situation awareness are overlapping con-structs that profit when team members have free cognitive resources. Each describes the awareness of a team member (or a whole team) about what is happening around him or her (Dickinson & McIntyre, 1997; Schulz, Endsley, Kochs, Gelb, & Wagner, 2013). Only if a team member knows what is going on can he or she anticipate and thus prepare for future actions. When an unexpected event occurs during an operation, an experienced nurse stays calm, knows what to do, anticipates future actions, and can provide direct help to the surgeon to help the patient before the surgeon has to give him or her an order or even explain what to do next.

These findings indicate that future studies should focus on the experience level of the team or specific team members. Experience can be investigated as an input factor influencing team processes or as a moderator of the process-outcome relationship. Some process behaviours might be effective for experienced teams but not for inexperienced ones. The converse is also likely. For example, explicit sharing of information could help an inexperienced team to develop a shared mental model of a situation and thereby increase per-formance but could distract an experienced team where all team members already have the information necessary to treat a patient and thus influencing performance negatively.

This thesis focuses on patient-related outcomes and omits the other part of output proposed by the team-performance framework: team outcomes. A whole separate area of research investigates the relationship between team processes and team outcomes such as burnout (Bobbio, Bellan, & Manganelli, 2012), job satisfaction (Bratt, Broome, Kelber, & Lostocco, 2000), or commitment (Lehmann-Willenbrock, Lei, & Kauffeld, 2012). Since these two outputs comprise two distinct strains of research on patient safety and occupational health, it is unsurprising that few studies investigate both types of outputs in relation with team processes. It might be interesting to see what the results would be if the same team processes were investigated in terms of both patient-related outputs and team outcomes or how these two outputs are related to each other.

Although Study A lists a number of articles investigating team processes and patient outcomes, overall this thesis focuses strongly on process performance, as shown in Studies B and D. Future studies should consider including both process and outcome performance measures.

4.2 Conclusion

This thesis makes a valuable contribution to the field of research on medical teams. Its theory-driven approach was chosen to address team processes and their related variables for an important reason: the nature of the process is always bounded and grounded by inputs and outputs in the organizational setting and cannot be seen as an independent entity operating in a vacuum (Ilgen, 1999).

As outlined in Chapter 2, there exist several good frameworks for addressing team behaviour in research and in practice. They form a solid basis for future research. In this field, the domain of theory is far more coherent than that of empirical studies, for several possible reasons. Theoretical frameworks addressing teams and team processes appear mainly in the psychological literature, whereas a large portion of empirical studies about team processes in healthcare is published in medical journals, with no reference to the psy-chological literature. The medical field has a more problem-driven and practical approach, whereas the psychological literature is more theory-driven. Theories and models fashioned for teams are indispensable for research and for practice, but often they are too complex because they cannot be tested statistically as a whole model or aid in solving concrete problems within a team in an organizational context (e.g. Burke, Stagl, Salas, Pierce, & Kendall 2006). A bal-ance needs to be found between theory-driven and problem-driven approaches.

Such a balance can only be achieved if researchers in the two domains collaborate with each other. The healthcare worker has a pragmatic understanding of how the teams work in his hospital, what tasks they do, and what problems they face. The responsibility of the organizational psychologist then is to study and listen to the people working in real environments and to use this knowledge to generate more cogent theories governing teams and team processes, ideally moving team research in industrial and organizational psychology forward into territory that simultaneously advances theory and practice.

5. Thesis Articles

The four studies (Study A-D) are presented as self-contained articles including their own structure and references. Language, formatting and citation style is based on the journal in which the article was published or to which it was submitted. All references from Chapter 1- 4 are listed at the very end of this thesis.

5.1 Study A: Do team processes really have an effect on team performance? A systematic literature review

Jan Schmutz[1] & Tanja Manser[1]

[1]Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Switzerland

Published under the following reference:

Schmutz, J. & Manser, T. (2013). Do team processes really have an effect on team performance? A systematic literature review. *British Journal of Anaesthesia, 110,* 529-544.

British Journal of Anaesthesia **110** (4): 529–44 (2013)
Advance Access publication 1 March 2013 · doi:10.1093/bja/aes513

BJA

Do team processes really have an effect on clinical performance? A systematic literature review

J. Schmutz* and T. Manser

Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Rue de Faucigny 2, 1700 Fribourg, Switzerland
* Corresponding author. E-mail: jan.schmutz@unifr.ch

Editor's key points

- This review has examined the impact of team process behaviours on clinical performance.
- Twenty-eight studies, which reported at least one relationship between team process or an intervention and outcome, were reviewed.
- Team process behaviours have been shown to influence performance.
- Training in team behaviours results in improved performance.

Summary. There is a growing literature on the relationship between team processes and clinical performance. The purpose of this review is to summarize these articles and examine the impact of team process behaviours on clinical performance. We conducted a literature search in five major databases. Inclusion criteria were: English peer-reviewed papers published between January 2001 and May 2012, which showed or tried to show (i) a statistical relationship of a team process variable and clinical performance or (ii) an improvement of a performance variable through a team process intervention. Study quality was assessed using predefined quality indicators. For every study, we calculated the relevant effect sizes. We included 28 studies in the review, seven of which were intervention studies. Every study reported at least one significant relationship between team processes or an intervention and performance. Also, some non-significant effects were reported. Most of the reported effect sizes were large or medium. The study quality ranged from medium to high. The studies are highly diverse regarding the specific team process behaviours investigated and also regarding the methods used. However, they suggest that team process behaviours do influence clinical performance and that training results in increased performance. Future research should rely on existing theoretical frameworks, valid, and reliable methods to assess processes such as teamwork or coordination and focus on the development of adequate tools to assess process performance, linking them with outcomes in the clinical setting.

Keywords: clinical competence; group processes; leadership; patient care team; patient safety

Breakdown in team processes such as coordination, leadership, or communication have frequently been associated with adverse events and patient harm[1–3] and the effectiveness of such team processes is central to the successful provision of patient care.[1 4 5] While recent reviews indicate that team processes are widely accepted as an important factor influencing clinical performance of medical teams,[1 5–8] a general framework is needed to classify and compare different studies on teamwork. In this review, we invoked McGrath's systemic input–process–output (IPO) framework[9] that has served as a foundation for numerous studies in team research[10–14] and has been adapted and used in clinical settings in recent years.[5 7 15–17]

According to this framework, *inputs* are preconditions influencing the processes in the team (e.g. team climate, task structure, leadership style). *Team processes* are defined as the cognitive, verbal, and behavioural activities going on while the team is working together (i.e. team communication, team leadership, team coordination, and team decision-making).[5 18 19] *Outputs* are the product of these processes. Either patient outcomes or team outcomes can be considered as *outputs* in a clinical setting.[5]

The IPO framework conceptualizes performance as an output that is directly influenced by team processes,[5 9] but does not provide explicit definitions of performance or a means by which to measure it. Various authors agree that there is both a process and an outcome-related aspect to performance.[20–22] The distinction between outcome and process performance measures is not always consistently used in the literature but should be borne in mind when aiming to establish an empirical evidence base on the relationships between team processes and outcomes.

Outcome performance measures such as mortality,[23] morbidity,[23] or length of stay[24] can be assessed objectively without consideration of the team process. Process performance measures, in contrast, are action-related aspects of performance embedded in the team processes.[15] Process performance measures are often more easily accessible and less influenced by other variables than outcome performance measures because they refer to directly observable behaviours executed by the team during patient treatment (e.g. measuring task execution time, rating specific behaviours according to medical guidelines).[25 26]

In the infancy of team research in medicine, the main aim was to generate a general understanding of which team processes influenced performance in which way. After qualitative studies investigating which team processes might be relevant to clinical performance,[27 28] quantitative studies were conducted to develop a clearer understanding of the impact of team processes on clinical performance. Studies investigated the association between team processes and either process performance[7 29] or outcome performance measures.[23] However, despite this improved understanding, it is still not clear how large the effect of these relationships is because in the majority of cases, no effect sizes are reported.

This systematic literature review aims to address this gap by analysing articles that investigate the relationship between team processes and clinical performance measures (i.e. process or outcome performance) and to report and compare the respective effect sizes. Furthermore, we will describe and discuss the different team processes and clinical performance measures used. This knowledge is needed to design targeted studies and effective interventions for patient care teams.

Methods

We conducted a literature search based on the recommendations of the PRISMA statement[30] consulting the databases PubMed, Science Direct, PsycINFO, PSYNDEXplus Literature, and Audiovisual Media. Additionally, a meta-search with Google Scholar was conducted; of which, only the first 50 results were examined. The search term used was PATIENT SAFETY combined with TEAMWORK, COMMUNICATION, or LEADERSHIP. In addition, a hand search was conducted based on the references of the identified articles. The literature search was conducted in May 2012.

Figure 1 provides an overview of the inclusion criteria and the five-step selection procedure. We selected English articles published in journals between January 2001 and May 2012 investigating the relationship between team processes and clinical performance. We selected articles that showed or tried to demonstrate (i) a statistical relationship between a team process variable and clinical performance (process or outcome performance) or (ii) an improvement of clinical performance (process or outcome performance) through an intervention targeting team processes.

We included only articles with performance measures. We excluded articles which used self-report data because surveys or interviews about the teams' own perception of performance can contain a self-report bias[31] and could potentially have distorted the results of this review. Intervention studies were only considered when targeting a team process behaviour (e.g. through training) and not implying structural changes (e.g. care pathways)[32] at the same time, because this would preclude distinguishing between effects of the training *vs* the structural change. We included studies using process or outcome performance measures. Since our main focus was on factors influencing patient care, we excluded studies measuring team outcomes (e.g. job satisfaction, stress, burnout).[5]

Each step was performed independently by two reviewers (J.S. and Mariel Dardel). The agreement was between 90% and 94% in each step. Any disagreement in the selection process was resolved by extensive discussion.

Rating of study quality

In order to assess the methodological quality of the selected articles, we used a rating system based mainly on the one proposed by Buckley and colleagues.[33] Since external validity is an important quality indicator, we replaced the single item by Buckley and colleagues with two items from a checklist by Downs and Black.[34] For intervention studies, three items concerning the quality of the intervention were added from Downs and Black. The question of triangulation was not applied to the intervention studies because the focus was on the effect of the intervention and we did not expect authors of intervention studies to triangulate multiple methods. The complete list and a detailed description of quality indicators can be found in Supplementary Table S3.

Each indicator was scored as '0' (not fulfilled), '0.5' (partially fulfilled), '1' (complete), or 'not mentioned' (i.e. information not explicitly provided and thus unclear whether the criterion has been fulfilled or not). Quality ratings were performed by J.S. A random sample of five studies was rated by T.M. We achieved consistency of 91%. Disagreements in the ratings were due to different interpretations of the descriptions in the articles and were resolved by discussion.

Data extraction

The following characteristics of the selected studies that were deemed most relevant were extracted, to evaluate the statistical relationships between team processes and clinical performance: team process behaviours, performance measures, participants, and results plus a description of the intervention in the case of intervention studies. Additionally, we calculated the effect size for every statistical process–output relationship reported in the selected studies based on the data provided in the articles. This enabled us to determine not only if team processes are significantly related with clinical performance but also how large this effect is and if it is large enough to be relevant for practical implications.[35] We report only significant and non-significant effects that were explicitly stated in the selected articles, although additional relationships may have been investigated but not reported.

Results

As can be seen from Figure 1, the initial search yielded 5383 articles. After excluding the irrelevant studies in stage 2, 887 articles remained. In stage 3, 784 studies were selected, of which 258 used quantitative methods and were retained for stage 4. After applying the final selection step, we identified 28 studies; of which, seven were intervention studies. Table 1 and Table 2 provide an overview of the relevant

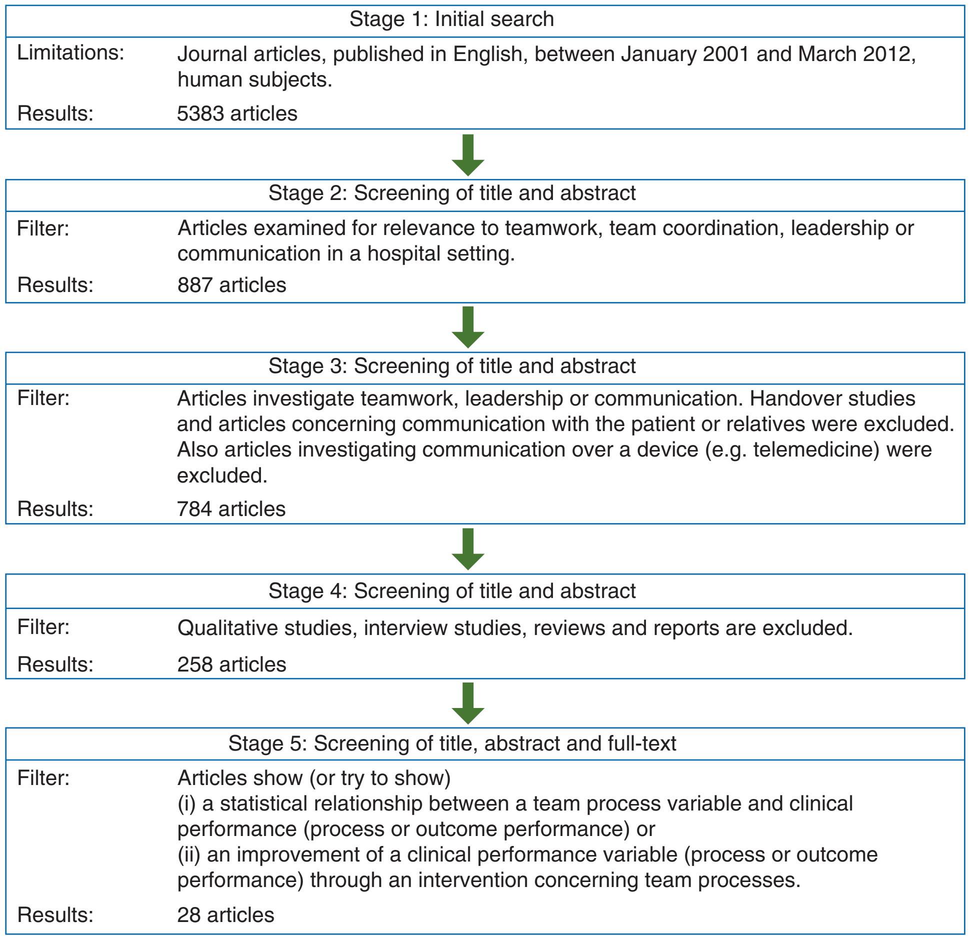

Fig 1 Systematic literature search, selection procedure and inclusion criteria.

characteristics pertaining to all the articles included in this review.

Team processes investigated and their measurement

The selected studies examined various team processes: communication,[23 24 36–38] coordination,[24 39–41] leadership,[7 24 31 42 43] non-technical skills,[29 44–49] team behaviour,[42] team monitoring behaviour,[50] and teamwork.[23 36 51] Six studies examined more than one team process behaviour.[23 24 31 36 42 43]

In reviewing the articles, we noted a high variability in the research approaches and measures used to study these team processes. As can be seen from Table 3, observational studies were most prominent. Most studies used video-based behaviour coding of data obtained in a simulator setting (n=10). Of the nine studies conducted in a clinical setting, three used video-based and six used live behaviour coding. Only three studies used surveys to collect team process data.

At the measurement instruments level, we found that four of the seven studies examining non-technical skills used the Surgical NOTECHS system.[29 44–46] The other three systems used were the Behavioural Marker Risk Index (BMRI),[47] the Anaesthetists' Non-Technical Skills (ANTS),[49] and one specific behavioural marker system for neonatal resuscitation.[48] Three of the six studies investigating communication used different observation systems[31 36 38] and the other three all used different questionnaires.[23 24 37] Three studies conceptualized the team processes under investigation as teamwork. Of these studies, one used the Safety Attitude Questionnaire (SAQ),[23] one used a rating system for teamwork behaviour,[51] and one study focused on events disrupting teamwork.[36] Of the five studies investigating leadership processes, four conducted observations but used different observation systems[31 42 43 52] and one study used a survey.[24] Of the four studies focusing on coordination, three[39–41] used the coding system of Manser and colleagues[53] and one assessed coordination using a survey.[24]

Process and outcome measures of clinical performance

Table 4 summarizes the 50 performance measures used in the 28 studies sorted into 41 process performance measures and nine outcome performance measures. Fourteen studies recorded deviations (i.e. errors, problems, or non-routine events during treatment) as a measure of process

Table 1 Characteristics of studies reporting relationships between team process behaviour and process or outcome performance

Study	Team process behaviour / research method and tool	Performance measure / method	Participants / setting	Results	Effect size*	Quality score (max=12)	'Not mentioned'§
Burtscher and colleagues[39]	*Team coordination* / behaviour coding of video data → coding system for coordination[53] consisting of 33 codes, which are grouped into five main categories: information management, task management, coordination via work environment, metacoordinaton, and other communication	Clinical performance of the anaesthesia induction / checklist-based rating system by experts	Anaesthesia staff, 19 anaesthetists, 14 nurses, teams of 2-4 persons / clinical setting (22 videos of routine anaesthesia inductions)	High performing teams show a more pronounced increase in task management in response to NRE in contrast to low-performing teams	Low performing teams: $\bar{x}_1\approx$ 23%† (routine) to $\bar{x}_2\approx$ 29%‡ (NRE) vs. high performing teams $\bar{x}_3\approx$ 16%† (routine) to $\bar{x}_4\approx$ 36 %‡ (NRE) (relative amount of time teams spent on task management); $t(20)=-2.75$, $p<.05$	10	1
Burtscher and colleagues[40]	*Adaptive coordination* while different phases of a treatment / behaviour coding of video data → coding system for coordination[53] consisting of 33 codes, which are grouped into five main categories: information management, task management, coordination via work environment, metacoordinaton, and other communication	Decision latency / time from the recognition of the asystole until the decision how to respond to it Execution latency / time from deciding what to do until restoration of sinus rhythm)	15 anaesthesia trainees, 15 anaesthesia nurses, teams of 2 persons / simulation (standard anaesthesia induction)	Negative association between decision latency and the anaesthesia trainees change in information management No association between other coordination aspects and decision latency or execution time	$r=-.49$ ($p=.003$) NS	9	1
Burtscher and colleagues[50]	*Team monitoring behaviour* / behaviour coding of video data → coding each time a team member was observing the action of a teammate	Clinical performance of the anaesthesia induction / checklist-based rating system by experts	31 anaesthesia resident, 31 anaesthesia nurses, teams of 2 persons / simulation (anaesthesia induction)	Negative association between team monitoring and performance	$r=-.44$ ($p=.02$)	10	1
Carlson and colleagues[42]	*Leadership* (LS) and *team behaviour* / behaviour coding of video data → global assessment of one dominant style of leadership (transactional LS, flexible/dynamic team LS, neither); rating (0–4) of four team behaviour categories (workload management, communication, prioritizing and reassessing priorities, vigilance)	Standard of care / expert assessment in consideration of behavioural guidelines (poor, marginal, standard of care)	113 3rd- year undergraduate medical students, teams of 2–3 persons / simulation (acute dyspnea)	Pos. association of the average team score (mean of the four dimensions) and standard of care No interrelation of LS style and standard of care	$r=.77$ ($p<.0001$) NS	8	2

Catchpole and colleagues[44]	*Non-technical skills* / behaviour coding of video and live data → Surgical NOTECHS measurement framework	Problems / observation Intraoperative performance / checklist-based rating system by experts Operating time	42 paediatric and orthopaedic operation teams / clinical setting (paediatric and orthopaedic operations)	Teams with effective teamwork have:		9.5	1
				Fewer minor problems per operation	† $\bar{x}_1 \approx 7.3$† (effective teams) vs. $\bar{x}_2 \approx 11.1$† (ineffective teams) $t=3.05$ $p=.004$		
				Higher intraoperative performance	† $\bar{x}_3 \approx 93.3\%$† (effective teams) vs. $\bar{x}_4 \approx 95.5\%$† (ineffective teams) $t=-3.25$ $p=.002$		
				Shorter operating times than teams with less effective teamwork	† $\bar{x}_5 \approx 195$min.† (effective teams) vs. $\bar{x}_6 \approx 153$min.† (ineffective teams) $t=2.25$ $p=.03$		
Catchpole and colleagues[73]	*Non-technical skills* / live behaviour coding → Surgical NOTECHS measurement framework	Operating time Errors in surgical technique / observation Other procedural problems and errors / observation	48 surgical operation teams / clinical setting (26 laparoscopic cholecystectomies, 22 carotid endarterectomies)	Association between situation awareness and errors in surgical technique	† ($F(2,42)=7.93$, $p=0.001$)	9.5	1
				Association between LS & Management and operating time	† ($F(2,42)=3.32$, $p = 0.046$)		
				Association between LS & Management score of the nurse and other procedural problems and errors	† ($F(5,1)=3.96$, $p=0.027$)		
				No association between other NOTECHS dimensions and any performance measure	NS		
Davenport and colleagues[23]	*Teamwork* and *communication* / survey → Safety attitudes questionnaire (SAQ)	Mortality (patient death in or out of the hospital 30 days after the operation) / data from the National Surgical Quality Improvement Program (NSQIP) Morbidity (patient having 1 or more postoperative complications up to 30 days after operation) / Data from the NSQIP)	6083 staff members of general and vascular surgery from 44 Veterans Affairs and 8 academic medical centres / clinical setting	Significant negative correlation between morbidity and		9	1
				Positive communication of surgical service care providers with attending doctors	$r=-.38$ ($p<0.01$)		
				Positive communication of surgical service care providers with residents	$r=-.25$ ($p<0.08$)		
				No interrelation of teamwork and mortality	NS		
				No interrelation of teamwork and morbidity	NS		

Continued

Table 1 *Continued*

Study	Team process behaviour / research method and tool	Performance measure / method	Participants / setting	Results	Effect size*	Quality score (max=12)	'Not mentioned'§
ElBardissi and colleagues[36]	*Teamwork* and *communication disruptions* / live behaviour coding → any occurrence concerning teamwork /communication that disrupted the flow of the operation	Errors (events failed its intended outcome) / observation	5 surgeons / clinical setting (31 cardiac surgical operations)	Positive association between teamwork disruptions and surgical errors	r=.67 (p<0.001)	11	0
Künzle and colleagues[7]	*Leadership* / behaviour coding of video data (structuring LS, content oriented LS and total amount of LS)	Execution time	12 anaesthesia teams / simulation (anaesthesia induction)	Negative association between execution time during routine and highly standardized phases and structuring LS and content oriented LS and total amount of LS No significant association between LS and execution time during a nonroutine event	 r=−.59 (p <.05) r=−.52 (p<.10) r=−.56 (p<.05) NS	9.5	1
Manojlovich and colleagues[37]	*Communication* / survey → ICU Nurse-Physician Questionnaire (4 scales: openness, accuracy, timeliness and understanding)	Ventilator-associated pneumonia / data from the hospital database Bloodstream infections / data from the hospital database Pressure ulcers / data from the hospital database	462 nurses from 25 ICUs / clinical setting	Negative association between timeliness and pressure ulcers. No significant association between overall communication or other subscales and outcome variables	 r=−.38 (p=.06) NS	9.5	1
Manser and colleagues[41]	*Team coordination* / behaviour coding of video data → coding system for coordination[53] consisting of 33 codes, which are grouped into five main categories: information management, task management, coordination via work environment, metacoordinaton, and other communication	Clinical performance / checklist-based rating system by experts	48 first year students, teams of 2 persons / simulation (malignant hyperthermia)	Positive association between task distribution and performance	r=−.466 (p<0.01)	10	1

Marsch and colleagues[43]	*Leadership, task distribution* and *information transfer* / behaviour coding of video data if the specific behaviour is present or not	Clinical performance / time-based scoring system for critical treatment steps	16 teams consisting of 2 nurses and 1 physician each / simulation (cardiopulmonary resuscitation)	Successful teams show more task distribution and	6/6 (successful teams showed task distribution) vs. 4/10 (failing teams showed task distribution), odds ratio can't be calculated	8.5	2
				more LS behaviour than failing teams.	*OR*=8 (successful teams show 8 times more likely LS behaviour than failing teams		
				No significant difference in information transfer	NS		
Mazzocco and colleagues[47]	*Non-technical skills* / live behaviour coding according to the behaviour marker risk index (BMRI)	Outcome score (1= no complications to 5=death or permanent disability) including complications and other significant postoperative outcomes / retrospective chart review of the concerning patients	130 physicians, nurses, operating room technicians, nurse anaesthetists / clinical setting (300 surgical cases)	Patients were more likely to experience death or a major complication when there were less teamwork behaviours	*OR*=4.82 (corrected for preoperative physical fitness)	10	1
Mishra and colleagues[29]	*Non-technical skills* / live behaviour coding → Surgical NOTECHS measurement framework	Technical errors / observation	Surgeons, anaesthetists, nurses / clinical setting (26 elective laparoscopic cholecystectomie operations)	Negative association between situation awareness of surgeons and technical errors	$\rho=-.718$ ($p<.001$)	10	1
Pollack and colleagues[24]	*Leadership, communication* and *coordination* / survey → organizational assessment tool	Mortality Bronchopulmonary dysplasia Periventricular / intraventricular haemorrhage or leukomalacia Retinopathy of prematurity Length of stay All outcomes were collected from clinical records	Nurses, physicians and respiratory therapists of 8 neonatal intensive care units / clinical setting (493 deliveries)	Positive association between leadership and PIVH/PVL	† $p<.001$	10.5	0
				Negative association between coordination and PIVH/PVL	† $p=.047$		
				No significant results for communication and the other outcome measures	NS		
Schraagen[74]	*Non-technical skills* / live behaviour coding → Surgical NOTECHS system	Non-routine events/ observation Operating time 30-day postsurgical outcome (uncomplicated, minor complication, major complications or death)	Paediatric cardiac surgical teams / clinical setting (40 operations)	Positive association between teamwork and operating times	$r=.45$ ($p<0.01$)	10	1
				No association between teamwork and non routine event	NS		
				No association between teamwork and outcome	NS		

Continued

Table 1 *Continued*

Study	Team process behaviour / research method and tool	Performance measure / method	Participants / setting	Results	Effect size*	Quality score (max=12)	'Not mentioned'§
Siassakos and colleagues[51]	*Teamwork* / behaviour coding of video data → Generic teamwork score (GTS)	Clinical efficiency score / check-list rating Time until turning the patient into the recovery position Time until administration of O_2 Time until venous blood sampling	24 teams consisting of 2 doctors and 4 midwives each / simulation (obstetric emergency)	Positive association between clinical efficiency score and GTS Negative association between GTS and time until turning the patient into the recovery position time until administration of O_2 time until venous blood sampling	r=.72¶ (p<0.001) r=−.38¶ (p=0.026) r=−.52¶ (p=0.002) r=−.60¶ (p<0.001)	9.5	1
Thomas and colleagues[48]	*Non-technical skills* / behaviour coding of video data → behaviour marker system	Compliance with Neonatal Resuscitation Program (NRP) guidelines / checklist-based rating system by experts	Neonatal resuscitation teams consisting of two providers, one physician one neonatal nurse / clinical setting (132 deliveries)	Negative correlation between total NRP noncompliance and Communication Management No correlation between leadership and total NRP noncompliance	ρ=−.021, p=0.014 ρ=−.020, p = 0.021 NS	10	1
Tschan and colleagues[31]	*Directive leadership* and *structuring inquiry* / behaviour coding of video data → coding system derived from the guidelines for cardiopulmonary resuscitation and other observational systems	Medical performance / checklist based rating of technical acts Percentage of time the patient did not show signs of normal circulation and received cardiovascular support	21 teams consisting of 3 nurses, 1 resident, 1 senior doctor / simulation (cardiac arrest)	Positive association between performance and directive LS of the first nurse and structuring inquiry of the nurses and directive LS of the resident (in the first 30s when he enters the room) and structuring inquiry of the senior physician	r=.445 (p<.05) r=.216 (p<.05) r=.522 (p<.01) r=.428 (p<.01)	10	1
Tschan and colleagues[38]	*Explicit reasoning* and *talking to the room* / behaviour coding of video data → sum of reasoning units, talking to the room present or not (dummy variable)	Diagnostic accuracy / evaluation of the team diagnosis	20 Groups consisting of 2 or 3 experienced physicians / simulation (anaphylactic shock)	Successful teams show more explicit reasoning (# of linked utterances) and more talking to the room than less successful teams No significant difference in the amount of information	† 4.0 (successful)>1.13 (successful with help)>1.0 (fail) $F(2,15)$=5.750; p = .014 $\hat{w}^2$=0.43 NS	10.5	1

				considered of successful and less successful teams			
Westli and colleagues[49]	*Non-technical skills / behaviour coding of video data* → ANTS system (revised)	-Performance score / checklist-based rating system by experts -Medical Management / overall rating from 1-5	27 trauma teams consisting of 1 surgeon 1 anaesthesiologist, 2 nurses, 1 radiographer / simulation (resuscitation)	Negative association between		9.5	1
				Medical Management and poor coordination	$r = -.36$ ($p < 0.01$)		
				Performance score and supporting behaviour	$r = -.37$ ($p < 0.01$)		
				Positive association between Medical Management and information exchange	$r = .34$ ($p < 0.01$)		
				No correlation between performance score and			
				coordination	NS		
				poor Coordination	NS		
				information exchange	NS		
				use of authority	NS		
				poor use of authority	NS		
				assessing capabiliteis	NS		
				poor supporting behaviour	NS		
				No correlation between Medical Management and			
				coordination	NS		
				use of authority	NS		
				poor use of authority	NS		
				assessing capabilities	NS		
				supporting behaviour	NS		
				poor supporting behaviour	NS		

* r, ρ and $\hat{w}^2$ effect sizes are interpreted as follows: r or $\rho = .10$ small effect; r or $\rho = .30$. medium effect; r or $\rho = .50$ large effect [62, 75]; $\hat{w}^2 = 0.01$ small effect, $\hat{w}^2 = 0.09$ medium effect, $\hat{w}^2 = 0.25$ large effect[73]
† The required information to calculate the effect sizes are not available. If available the absolute sizes are indicated instead.
‡ Means are assessed out of figures. The exact means are not mentioned in the text.
¶ Kendall's Tau (τ) was transformed into r according to Walker[74]
§ 'Not mentioned' means it was unclear if something has been done or not based on the information provided in the article. NS, Not Significance.

Table 2 Characteristics of team process behaviour interventions and their impact on performance. Team process measures used in the intervention studies are not listed here because for these studies, the focus is on the effect of the intervention on performance and not on the process. *The required information to calculate the effect sizes was not available. The absolute sizes are indicated instead. †$\hat{w}^2$ effect sizes are interpreted as follows: $\hat{w}^2 \geq 0.01$ small, $\hat{w}^2 \geq 0.09$ medium, and $\hat{w}^2 \geq 0.25$ large;[71] Cohen's *d* effect sizes are interpreted as follows: $d \geq 0.20$ small, $d \geq 0.50$ medium, and $d \geq 0.80$ large.[71] ‡'Not mentioned' means it was unclear if something has been done or not based on the information provided in the article

Study	Intervention/design/team process measure*	Performance measure/ method	Participants	Results	Effect size†	Quality score (max=14)	'Not mentioned'‡
Fernandez Castelao and colleagues[60]	Video based crew resource management training/quasi-experimental control group post-test design/no team process measure	No-flow time (time with no chest compression)	Four-person medical student teams, 26 teams in the experimental group, 18 teams in the control group	Less no flow time in the post-intervention group comparing with the control group	$\bar{x}_1 = 36.3\%$ (control) *vs* $\bar{x}_2 = 31.4$ (experimental) (P=0.014)*	11	0
Kalisch and colleagues[54]	Staff teamwork and engagement enhancement intervention/quasi-experimental uncontrolled pre-test–post-test design/post-interview about teamwork	Fall rates per 1000 patient days/information from patient report	49 nurses, six unit secretaries of a community hospital	Patient fall rates decreased after the intervention	$\bar{x}_1 = 7.73$ to $\bar{x}_2 = 2.99$ falls per 1000 patient days (t=3.98, P<0.001)*	9	2
McCulloch and colleagues[55]	Intervention based on principles of civil aviation crew resource management/quasi-experimental uncontrolled pre-test–post-test design/NOTECHS and SAQ teamwork climate score	Operating technical errors/ observation Operating time Length of stay	Surgeons, anaesthetists, nurses performing 48 operations in the pre-intervention group and 55 operations in the post-intervention group	Less operating technical errors after the intervention No reduction in operating time No reduction in length of stay	d=0.63 (P=0.009) NS NS	10.5	0
Morey and colleagues[57]	Emergency Team Coordination Course (ETCC)/quasi-experimental control group design with one pre- and two post-tests/Behaviour Anchored Rating Scales (BARS)	Errors/observation	Physicians, nurses, and technicians of six emergency departments (EDs) in the experimental group (n=684) and three EDs in the control group (n=374)	Decrease in the clinical error rate in the post-intervention group No significant difference between the experimental and control group	d=1.93 (P=0.039) NS	12	1
Nielsen and colleagues[58]	MedTeams Labor and Delivery Team Coordination Course based on crew resource management trainings/cluster-randomized control group design with no pre-test/no team process measure	Adverse maternal Outcome Index (number of patients with one or more adverse outcomes divided by the total number of deliveries)/information from patient report	Obstetrician, anaesthesiologist, and nurses of seven hospitals (obstetrics) in the experimental group (n=1307) and eight hospitals in the control group	Significant reduction in Caesarean delivery decision to incision	$\bar{x}_1 = 33.3$ min (control) *vs* $\bar{x}_2 = 21.2$ (experimental) (P=0.039)*	12	0

		Immediate Caesarean delivery decision to incision (min)/ observation		No reduction in adverse maternal outcomes in the experimental group	NS		
Phipps and colleagues[59]	Adaptation of a crew resource management training with a simulator/quasi-experimental uncontrolled pre-test–post-test design/safety culture survey from the Agency for Healthcare Research and Quality (AHRQ) consisting a teamwork and communication scale	Adverse Outcomes Index (AOI)/report	Nurse midwives, gynaecologist physicians, residents, anaesthetists, and unit secretaries of a labour and delivery unit with ~9200 births per year	Reduction in adverse events in the post-CRM group	AOI index decreases significantly from $\bar{x}_1=0.052$ to $\bar{x}_2=0.043$ (no P-value stated)*	9.5	1
Wolf and colleagues[56]	Medical team training (CRM based)/quasi-experimental uncontrolled design with one pre- and two post-tests/SAQ (teamwork climate scale)	Case delays/information from patient report	Operating teams consisting of surgeons, residents, anaesthesiologists, and nurses of a hospital with eight operating theatres	Reduction in cases with any delay in the post-intervention group	Cases with any delay significantly decreased from $\bar{x}_1=23.2\%$ to $\bar{x}_2=10\%$ ($P<0.0001$)*	11	1

Table 3 Methods and research settings for studying the team process–performance relationship

Methods/settings	Number of articles	Study reference number
Live observation/clinical setting	6	29,36,44,47,45,72
Video-based observation/clinical setting	3	39,44,48
Video-based observation/simulation	10	7,31,38,40–43,49–51
Survey/clinical setting	3	23,24,37

performance. Performance checklists based on clinical guidelines were the next frequently used performance measure (n=10) followed by the time until a specific treatment is conducted (n=7). The outcome performance measure used most frequently was complications after treatment (n=4).

Effects of team processes on performance in the non-intervention studies

In total, the 21 studies reported 66 relationships of a team process variable with a performance variable. Forty of these effects were significant and 26 were non-significant. Thirteen of the 21 non-intervention studies calculated correlations to investigate this relationship. More than one performance measure was used by 15 studies and 12 of these reported both non-significant and significant effects. Only six studies investigated just one effect and assessed only one performance measure. All of them were significant.

No study explicitly reported effect sizes. The effect sizes calculated are shown in Table 1 and Table 2. They range from very high (r=0.77)[42] to small ($\rho=-0.02$).[48] Only one study reported a small effect,[48] while all the others described effects considered as large or medium.

Interventions targeting team process behaviours

The interventions were carried out in community hospitals,[54] operating theatres,[55 56] emergency departments,[57] and labour and delivery units.[58 59] Five of the seven intervention studies used training explicitly based on crew resource management (CRM) principles,[55 56 58–60] while the other two studies included some CRM elements such as an introduction to teamwork and non-technical skills. According to the brief descriptions in the articles, it appears that all interventions were of similar content. Typical topics discussed in the training were principles of teamwork and human factors, situation awareness, improvement of team skills, communication, and leadership. The duration of the training ranged from 1 to 2 days and included methods such as theoretical lectures on CRM principles, video analysis, and role-playing. Unfortunately, an exact comparison of the interventions is not possible due to the limited descriptions of the training provided in the articles. Table 2 summarizes the effects of the seven interventions, all

Table 4 Performance measures used. *If multiple performance measures are used in one article, the study is mentioned several times. NREs, non-routine events

Performances measures used	Total number of performance measures used	Study reference number*
Process performance measure		
Deviations (errors, problems, NREs during the treatment)	14	24,24,24,29,36,37,37,37,44,44,55,57,45,72
Case delays	1	56
Length of stay	2	24,55
Operating time	5	7,44,55,45,72
Percentage of time the patient receives a specific treatment	2	31,60
Time until a specific treatment is conducted	7	40,40,43,51,51,51,58
Performance checklists	10	31,39,41,42,44,48,49,49,50,51
Outcome performance measures		
Complications after operation	4	47,58,59,72
Diagnostic accuracy	1	38
Fall rates	1	54
Morbidity	1	23
Mortality	2	23,24

indicating significant improvements of performance after the intervention. The intervention studies reported 11 effects on a performance measure; of which, seven were significant. Three studies assessed more than one performance measure. Only two studies indicated all the information to calculate the effect size and they reported one medium[55] and one large effect.[57]

Quality of the selected studies

A complete list of the quality ratings for every article can be found in the Supplementary Table S1 and Supplementary Table S2. The study quality ratings ranged from 9 to 12 points out of 14 for the intervention studies and from 8 to 11 out of 12 points for the other studies. Overall, data collection methods were found to be reliable and valid to answer the specific research questions. Two common problems were the poor discussion of potential confounding factors and the use of a single data collection method instead of strengthening the results through triangulation.

All non-intervention studies were prospective. In general, research questions were clearly stated, methods well described, analyses were appropriate, and the conclusions clearly justified by the results.

All intervention studies used quasi-experimental or clustered designs. Only three of the seven intervention studies applied a control group design, while the other four were pre-test–post-test studies. Two studies included a follow-up post-test to investigate long-term effects. All intervention studies provided unspecific descriptions of the conducted interventions limiting their reproducibility.

Other study characteristics

The studies included participants of various professions examining teams consisting of anaesthetists, nurses, medical students, paediatricians, surgeons, operating theatre technicians, and midwives. In four studies, the participants were uniprofessional.[37 38 41 42]

Discussion

The aim of our systematic literature review was to consolidate the statistical evidence for the effects of team processes on clinical performance in patient care teams. Furthermore, we provide an overview of all team process and performance measures used in these studies that will inform future research in this field regarding the strength and weaknesses of current measures and necessary developments.

Focusing on the process–performance relationship, this review found that significant progress has been made in recent years. Most studies report strong effects indicating that team processes are significantly influencing clinical performance. However, we identified areas for improvement with regard to defining and measuring both team processes and clinical performance. Our systematic analysis of study quality also points at possible improvements in both study design and reporting.

Most studies did not refer to a conceptual framework. They sometimes used vague definitions of the two concepts 'team process behaviours' and 'performance' and a broad range of measurement approaches was also seen. An appropriate scientific definition and explicit reference to a common conceptual framework are prerequisites for comparing studies that investigate a broad spectrum of team process behaviours. Such a framework aids in study design and interpretation of results. Although the IPO model is rather simple, it is widely accepted and has proven useful in various teamwork settings. The IPO model facilitates the research process by providing a clear structure of potential relationships upon which to focus (e.g. the impact of team mental models as an input on team process behaviours such as decision-making or the relationship between leadership processes within the team and subjective outcomes such as staff well-being).

While more complex models such as the input–throughput–output model of team adaptation of Burke and colleagues[61] have been developed to reflect the complexity of teamwork, these models are often too complex for isolating research questions that can be tested in an actual work setting. We have to strive for a balance between complexity and feasibility for these models to be useful in guiding team research in healthcare and in conceptually clarifying the relevant inputs, team process behaviours, and outcomes.

Most studies measuring team process behaviours have used observational methods. This is a more time-consuming method than questionnaire-based designs, but generally, observational methods are the most appropriate way to describe and measure processes. It avoids the problems of subjectivity and recall bias inherent in questionnaire-based designs,[62 63] especially in stressful situations. While questionnaire-based assessment provides a more general picture of team members' perceptions of team processes, observation methods capture the actions actually performed by the team members.

Moreover, to assure a valid assessment of team processes, observation systems should be as holistic and detailed as possible instead of focusing on a single behavioural facet. The observation system should allow for categorizing all behaviours performed by the team to investigate the interactions between different team behaviours and their relative contribution to the outcome.

The two observation systems used most frequently in the selected studies was the observation method of Manser and colleagues[53] and the behavioural marker system Surgical NOTECHS.[44] The system of Manser and colleagues assesses different aspects of team coordination including information management, task management, coordination via work environment, and others. The NOTECHS system includes behavioural dimensions such as leadership, teamwork, problem-solving, situation awareness, etc. The difference between these two systems is that the former is descriptive, that is, it objectively records actions of the team continuously without any evaluation. Other authors also use descriptive, non-evaluative systems.[31 38 43 52] In the Surgical NOTECHS system, the target behaviours are rated on a scale from 1 to 4 for a defined teamwork episode (e.g. anaesthesia induction). This evaluative component may artificially increase the relationship with performance ratings, while descriptive observation systems provide more objective data on the team process. Thus, it is critical to define performance measures that are truly independent of the team process measures.

The ultimate outcome of high performance in healthcare should be patient safety. As patient safety itself is difficult to measure and to relate to specific team process, various proxy measures have been used. The studies included in this review used many different measures to assess clinical performance that can be grouped into process performance measures and outcome performance measures.

Outcome performance measures are related to the result of the actions and depend on more than just individuals' behaviour.[21] For example, it is known from resuscitation that the duration of a patient's arrest, the primary arrhythmia, and patient age are better predictors for survival than the actual performance of the clinicians performing the resuscitation.[64] In clinical settings, it is impossible to take in to account all the factors potentially influencing performance, but there are ways to control some of them. For example, the ASA patient classification index has been used to classify patients' risk for complications taking into account the history of the patient[47] and the score for neonatal acute physiology (SNAP) has been used to assess the possibility of complications accounting for the newborn's physiology.[24] Another way to control or balance for confounders are large sample sizes that are often not feasible for very detailed, resource-intensive analyses of team processes and sometimes difficult to obtain in healthcare; especially in field studies requiring a high number of specific, comparable cases performed by care providers with predefined experience levels. In addition, ethical issues sometimes limit the spectrum of cases that can be studied using live observation in clinical settings.

Besides outcome performance measures, the processes leading to this outcome are also good indicators for performance (e.g. timely start of the correct treatment for the patient). These process performance measures refer to what an individual does in a specific work situation and are therefore less influenced by other factors.[21] Process performance can be assessed in almost every setting. During simulation, where it is hardly possible to assess patient-related outcomes, process performance measures are preferable. Performance checklists, for example, that take into account the most important actions for a specific treatment and evaluate those across the whole process provide a valid and reliable method to assess process performance if developed systematically. This includes a theoretical foundation and an integration of official guidelines and experiences of several experts[65] (e.g. through a Delphi process as, for example, done by Burtscher and colleagues).[39 66]

For intervention studies, the results of our review showed that training targeting team process behaviours do influence various outcomes. All the interventions focused exclusively on outcome performance measures. Therefore, one can only assume that the interventions influenced the team processes, which in turn led to better outcomes. This assumption will require further empirical testing to improve our understanding of the mechanisms through which the improvements have been achieved. Unfortunately, no effect sizes could be calculated for most studies, so it is difficult to determine how strong these effects really are. Also, each study referred to a different intervention, none of which was sufficiently described to be reproducible (for a discussion of this issue, see also Buljac-Samardzic and colleagues).[67]

Limitations

Several limitations of this systematic review have to be taken into account when interpreting the results. We focused only on English, peer-reviewed articles and did not include books or grey literature, so we may have missed relevant publications. Owing to the difficulties with publishing non-significant results,[68] there may be other studies which found no effect of team process behaviour or interventions on performance which we could not access.

In this review, we listed the team processes as they appear in the selected articles. However, if two studies used the same term, this does not necessarily mean they also referred to the same definition of this team process. Furthermore, we focused exclusively on the relationship between team processes and outputs. However, we acknowledge that team processes are not independent of input factors. Specific input factors could neutralize the relationship between processes and outputs. For example, Burtscher and colleagues[50] found a relationship between team monitoring behaviour and performance only when the team members had a shared mental model of the task.

Future research

This review identified some gaps in the literature on the relationship between team process behaviours and clinical performance.

Since most studies focus on acute patient care, more research needs to be done in other domains of healthcare such as long-term care. Also, only two studies included in this review conducted a follow-up post-test to check if the interventions also had a long-term effect. Thus, studies investigating team processes using a longitudinal design are needed; especially for intervention studies.

In comparison with the sizable literature on the importance of team process behaviour in healthcare, little research has actually investigated the statistical effects on process or outcome performance. To achieve this, valid process performance measures are required and will have to be developed systematically. That is, the relationship of process performance (e.g. checklist-based assessments) and outcomes has to be tested in controlled clinical studies to assure their validity and reliability for assessing performance in clinical and simulated settings.

Of course, there is no single best performance measure. In occupational psychology, it is widely accepted that performance is a multidimensional construct.[15 21] Thus, to get an accurate picture of performance, future studies should use multiple process performance measures or even combinations of process and outcome performance measures.[24 37 44 45]

To further our understanding of specific team processes such as coordination or leadership studies using the same observation systems and performance measures are needed. We gave a brief overview including pros and cons of different measurement methods and future research should take these considerations into account. This will result in more conceptual and methodological consistency and more definitive findings about the effects of team process behaviours on performance (e.g. supported by meta-analyses).

Our results suggest that team processes in general are clinically relevant because they have an effect on patient outcomes. A large effect size is an indicator for high clinical relevance; however, they are not necessarily linked.[69] For a more precise assessment of clinical relevance, future research should include other factors than statistical results as well.

Some studies included in this review show rather small or no correlations between team processes and performance.[48 49] It is not certain if this is due to unclear or inconsistent definitions of the constructs, validity issues, or confounders. However, we are sure that future research will help to explain and clarify these contradictory results with (i) clear and consistent definitions of the team processes investigated and (ii) more complete descriptions of the mechanisms linking specific team processes to specific performance measures that is embedded in a theoretical framework. Lingard and colleagues[70] illustrate how this could be done using the example of communication patterns related to collaborative work processes and patient safety. In this way, future research will deliver a more accurate picture of the relationship between team processes and performance. With this knowledge, we will be able to design more effective and successful team interventions and implementation strategies which will help to improve patient safety.

Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

Acknowledgements

We are grateful to Mariel Dardel for screening the literature and her help in the selection process of relevant articles.

Declaration of interest

None declared.

Funding

This work was supported by the Swiss National Science Foundation (grant number, PP00P1_128616).

References

1 Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiol Scand* 2009; **53**: 143–51

2 Manojlovich M, DeCicco B. Healthy work environments, nurse–physician communication, and patients outcomes. *Am J Crit Care* 2007; **16**: 536–43

3 Kohn LT, Corrigan J, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press, 1999

4 Salas E, Rosen MA, King H. Managing teams managing crises: principles of teamwork to improve patient safety in the emergency room and beyond. *Theor Issues Ergon Sci* 2007; **8**: 381–94

5 Reader TW, Flin R, Mearns K, Cuthbertson BH. Developing a team performance framework for the intensive care unit. *Crit Care Med* 2009; **37**: 1787–93
6 Martin JS, Ummenhofer W, Manser T, Spirig R. Interprofessional collaboration among nurses and physicians: making a difference in patient outcome. *Swiss Med Wkly* 2010; **140**: w13062
7 Künzle B, Kolbe M, Grote G. Ensuring patient safety through effective leadership behaviour: a literature review. *Safety Sci* 2010; **48**: 1–17
8 Rutherford J, Flin R, Mitchell L. Non-technical skills of anaesthetic assistants in the perioperative period: a literature review. *Br J Anaesth* 2012; **109**: 27–31
9 McGrath JE. *Social Psychology: A Brief Introduction*. Holt, New York: Rinehart and Winston, 1964
10 Gladstein DL. Groups in context: a model of task group effectiveness. *Adm Sci Quart* 1984; **29**: 499–517
11 Ilgen DR, Hollenbeck JR, Johnson M, Jundt D. Teams in organizations: from input–process–output models to IMOI models. *Annu Rev Psychol* 2005; **56**: 517–43
12 Hackman JR, Morris CG. Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. In: Berkowitz L, ed. *Advances in Experimental Social Psychology*. New York: Academic Press, 1975; 45–99
13 Hackman JR. The design of work teams. In: Lorsch JW, ed. *Handbook of Organizational Behavior*. Englewood Cliffs, NJ: Prentice-Hall, 1987; 315–42
14 Jarboe S. A comparison of input–output, process–output, and input–process–output models of small group problem-solving effectiveness. *Commun Monogr* 1988; **55**: 121–42
15 Tschan F, Semmer N, Vetterli M, Gurtner A, Hunziker S, Marsch SU. Developing observational categories for group process research based on task and coordination requirement analysis: examples from research on medical emergency-driven teams. In: Boos M, Kolbe M, Kappeler P, Ellwart T, eds. *Coordination in Human and Primate Groups*. Berlin: Springer, 2011; 93–115
16 Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA. The influence of shared mental models on team process and performance. *J Appl Psychol* 2000; **85**: 273–83
17 Flin R. Identifying and training non-technical skills for teams in acute medicine. *Qual Saf Health Care* 2004; **13**: i80–4
18 Campion MA, Medsker GJ, Higgs AC. Relations between work group characteristics and effectiveness: implications for designing effective work groups. *Person Psychol* 1993; **46**: 823–50
19 Marks MA, Mathieu JE, Zaccaro S. A temporally based framework and taxonomy of team processes. *Acad Manage Rev* 2001; **26**: 356–76
20 Anderson N, Ones DS, Sinangil HK, Viswesvaran C. *Handbook of Industrial, Work and Organizational Psychology: Personnel Psychology*. London: Sage Publications Ltd, 2001
21 Sonnentag S, Frese M. Performance concepts and performance theory. In: Sonnentag S, ed. *Psychological Management of Individual Performance*. Chichester: John Wiley & Sons, Ltd, 2002; 1–25
22 Nerdinger FW, Blickle G, Schaper N. *Arbeits-und Organisationspsychologie*. Berlin: Springer, 2011
23 Davenport D, Henderson W, Mosca C, Khuri S, Mentzerjr R. Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J Am Coll Surg* 2007; **205**: 778–84
24 Pollack MM, Koch MA. Association of outcomes with organizational characteristics of neonatal intensive care units. *Crit Care Med* 2003; **31**: 1620–9
25 Adler MD, Vozenilek JA, Trainor JL, *et al.* Comparison of checklist and anchored global rating instruments for performance rating of simulated pediatric emergencies. *Simul Healthc* 2010; **6**: 18–24
26 Donoghue A, Nishisakia A, Suttona R, Halesc R, Boulet JR. Reliability and validity of a scoring instrument for clinical performance during Pediatric Advanced Life Support simulation scenarios. *Resuscitation* 2010; **81**: 331–6
27 Wiles R, Robison J. Teamwork in primary care: the views and experiences of nurses, midwives and health visitors. *J Adv Nurs* 1994; **20**: 324–30
28 Körner M, Bengel J. Teamarbeit und Teamerfolg bei multi-und interdisziplinären Teams in der medizinischen Rehabilitation teamwork and team success in multidisciplinary and interdisciplinary teams in medical rehabilitation. *Rehabilitation* 2004; **43**: 348–57
29 Mishra A, Catchpole K, Dale T, McCulloch P. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc* 2007; **22**: 68–73
30 Liberati A, Altman DG, Tetzlaff J, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009; **6**: e1000100
31 Tschan F, Semmer NK, Gautschi D, Hunziker P, Spychiger M, Marsch SU. Leading to recovery: group performance and coordinative activities in medical emergency driven groups. *Hum Perform* 2006; **19**: 277–304
32 Thomas EJ. Improving teamwork in healthcare: current approaches and the path forward. *Br Med J* 2011; **20**: 647–50
33 Buckley S, Coleman J, Davison I, *et al.* The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11. *Med Teach* 2009; **31**: 282–98
34 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; **52**: 377–84
35 Kirk RE. Promoting good statistical practices: some suggestions. *Educ Psychol Meas* 2001; **61**: 213–8
36 ElBardissi AW, Wiegmann DA, Henrickson S, Wadhera R, Sundt Iii TM. Identifying methods to improve heart surgery: an operative approach and strategy for implementation on an organizational level. *Eur J Cardiothorac Surg* 2008; **34**: 1027–33
37 Manojlovich M, Antonakos CL, Ronis DL. Intensive care units, communication between nurses and physicians, and patients' outcomes. *Am J Crit Care* 2009; **18**: 21–30
38 Tschan F, Semmer NK, Gurtner A, *et al.* Explicit reasoning, confirmation bias, and illusory transactive memory: a simulation study of group medical decision making. *Small Group Res* 2009; **40**: 271–300
39 Burtscher MJ, Wacker J, Grote G, Manser T. Managing nonroutine events in anesthesia: the role of adaptive coordination. *Hum Fact* 2010; **52**: 282–94
40 Burtscher MJ, Manser T, Kolbe M, *et al.* Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. *Br J Anaesth* 2011; **106**: 1–6
41 Manser T, Harrison TK, Gaba DM, Howard SK. Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesth Analg* 2009; **108**: 1606–15
42 Carlson J, Min E, Bridges D. The impact of leadership and team behavior on standard of care delivered during human patient

simulation: a pilot study for undergraduate medical students. *Teach Learn Med* 2009; **21**: 24–32

43 Marsch SCU, Müller C, Marquardt K, Conrad G, Tschan F, Hunziker PR. Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. *Resuscitation* 2004; **60**: 51–6

44 Catchpole KR, Giddings AE, Wilkinson M, Hirst G, Dale T, de Leval MR. Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007; **142**: 102–10

45 Catchpole K, Mishra A, Handa A, McCulloch P. Teamwork and error in the operating room: analysis of skills and roles. *Ann Surg* 2008; **247**: 699–706

46 Schraagen JM, Schouten T, Smit M, *et al.* A prospective study of paediatric cardiac surgical microsystems: assessing the relationships between non-routine events, teamwork and patient outcomes. *Br Med J* 2011; **20**: 599–603

47 Mazzocco K, Petitti DB, Fong KT, *et al.* Surgical team behaviors and patient outcomes. *Am J Surg* 2009; **197**: 678–85

48 Thomas E, Sexton J, Lasky R, Helmreich R, Crandell D, Tyson J. Teamwork and quality during neonatal care in the delivery room. *J Perinatol* 2006; **26**: 163–9

49 Westli H, Johnsen B, Eid J, Rasten I, Brattebø G. Teamwork skills, shared mental models, and performance in simulated trauma teams: an independent group design. *Scand J Trauma Resusc Emerg Med* 2010; **18**: 47

50 Burtscher MJ, Kolbe M, Wacker J, Manser T. Interactions of team mental models and monitoring behaviors predict team performance in simulated anesthesia inductions. *J Exp Psychol Appl* 2011; **17**: 257

51 Siassakos D, Fox R, Crofts JF, Hunt LP, Winter C, Draycott TJ. The management of a simulated emergency: better teamwork, better performance. *Resuscitation* 2011; **82**: 203–6

52 Künzle B, Zala-Mezo E, Kolbe M, Wacker J, Grote G. Substitutes for leadership in anaesthesia teams and their impact on leadership effectiveness. *Eur J Work Organ Psychol* 2010; **19**: 505–31

53 Manser T, Howard SK, Gaba DM. Adaptive coordination in cardiac anaesthesia: a study of situational changes in coordination patterns using a new observation system. *Ergonomics* 2008; **51**: 1153–78

54 Kalisch BJ, Curley M, Stefanov S. An intervention to enhance nursing staff teamwork and engagement. *J Nurs Adm* 2007; **37**: 77–84

55 McCulloch P, Mishra A, Handa A, Dale T, Hirst G, Catchpole K. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care* 2009; **18**: 109–15

56 Wolf FA, Way LW, Stewart L. The efficacy of medical team training: improved team performance and decreased operating room delays: a detailed analysis of 4863 cases. *Ann Surg* 2010; **252**: 477–83

57 Morey JC, Simon R, Jay GD, *et al.* Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002; **37**: 1553–81

58 Nielsen PE, Goldman MB, Mann S, *et al.* Effects of teamwork training on adverse outcomes and process of care in labor and delivery: a randomized controlled trial. *Obstet Gynecol* 2007; **109**: 48–55

59 Phipps MG, Lindquist DG, McConaughey E, O'Brien JA, Raker CA, Paglia MJ. Outcomes from a labor and delivery team training program with simulation component. *Am J Obstet Gynecol* 2011; **206**: 1–7

60 Fernandez Castelao E, Russo SG, Cremer S, *et al.* Positive impact of crisis resource management training on no-flow time and team member verbalisations during simulated cardiopulmonary resuscitation: a randomised controlled trial. *Resuscitation* 2011; **82**: 1338–43

61 Burke CS, Stagl KC, Salas E, Pierce L, Kendall D. Understanding team adaptation: a conceptual analysis and model. *J Appl Psychol* 2006; **91**: 1189–207

62 Bortz J, Döring N. *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler*. Berlin: Springer Verlag, 2006

63 Greve W, Wentura D. *Wissenschaftliche Beobachtung: Eine Einführung*. Weinheim: BeltzPVU, 1997

64 Cooper S, Cade J. Predicting survival, in-hospital cardiac arrests: resuscitation survival variables and training effectiveness. *Resuscitation* 1997; **35**: 17–22

65 Landon BE, Normand SLT, Blumenthal D, Daley J. Physician clinical performance assessment. *J Am Med Assoc* 2003; **290**: 1183–9

66 Gordon TJ. *The Delphi Method.* In: Glenn JC, ed. Futures Research methodology series. Washington, D. C.: United Nations University; 1994

67 Buljac-Samardzic M, Dekker-van Doorn CM, Van Wijngaarden JDH, Van Wijk KP. Interventions to improve team effectiveness: a systematic review. *Health Policy* 2010; **94**: 183–95

68 Easterbrook PJ, Gopalan R, Berlin J, Matthews DR. Publication bias in clinical research. *Lancet* 1991; **337**: 867–72

69 Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: history, application, and current practice. *Clin Psychol Rev* 2001; **21**: 421–46

70 Lingard L, Whyte S, Espin S, Ross Baker G, Orser B, Doran D. Towards safer interprofessional communication: constructing a model of 'utility' from preoperative team briefings. *J Interprof Care* 2006; **20**: 471–83

71 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum, 1988

72 Schraagen JM. Dealing with unforeseen complexity in the OR: the role of heedful interrelating in medical teams. *Theor Issues Ergon Sci* 2011; **12**: 256–72

73 Rasch B, Friese M, Hofmann W, Naumann E. *Quantitative Methoden: Einfuhrung in die Statistik 1*. 2 Edn. Heidelberg: Springer Medizin Verlag, 2006

74 Walker DA. JMASM9: converting Kendall's tau for correlational or meta-analytic analyses. *J Mod Appl Stat Meth* 2003; **2**: 525–30

Handling editor: R. P. Mahajan

5.2 Study B: Five steps to develop checklists for evaluating clinical performance: An integrative approach

Jan Schmutz[1], Walter J. Eppich[2], Florian Hoffmann[3], Ellen Heimberg[4], Tanja Manser[1]

[1]Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Fribourg, Switzerland.
[2] Northwestern University Feinberg School of Medicine, Chicago, Illinois.
[3] Dr. von Hauner University Children‘s Hospital, Munich, Germany.
[4] University Children‘s Hospital, Tübingen, Germany.

Published under the following reference:

Schmutz, J., Eppich, W. J., Hoffmann, F., Heimberg, E., & Manser, T. (2014). Five steps to develop checklists for evaluating clinical performance: An integrative approach. *Academic Medicine, 89.*

LWW | 04/17/14 | 4 Color Fig(s):0 | 10:3 | Art: ACADMED-D-13-01053

Five Steps to Develop Checklists for Evaluating Clinical Performance: An Integrative Approach

Jan Schmutz, MSc, Walter J. Eppich, MD, MEd, Florian Hoffmann, MD,
Ellen Heimberg, MD, and Tanja Manser, PhD

Abstract

Purpose

The process of developing checklists to rate clinical performance is essential for ensuring their quality; thus, the authors applied an integrative approach for designing checklists that evaluate clinical performance.

Method

The approach consisted of five predefined steps (taken 2012–2013). *Step 1:* On the basis of the relevant literature and their clinical experience, the authors drafted a preliminary checklist. *Step 2:* The authors sent the draft checklist to five experts who reviewed it using an adapted Delphi technique. *Step 3:* The authors devised three scoring categories for items after pilot testing. *Step 4:* To ensure that the changes made after pilot testing were valid, the checklist was submitted to an additional Delphi review round. *Step 5:* To weight items needed for accurate performance assessment, 10 pediatricians rated all checklist items in terms of their importance on a scale from 1 (not important) to 5 (essential).

Results

The authors have illustrated their approach using the example of a checklist for a simulation scenario of septic shock in an infant. The five-step approach resulted in a valid, reliable tool and proved to be an effective method to design evaluation checklists. It resulted in 33 items, most consisting of three scoring categories.

Conclusions

This approach integrates published evidence and the knowledge of domain experts. A robust development process is a necessary prerequisite of valid performance checklists. Establishing a widely recognized standard for developing evaluation checklists will likely support the design of appropriate measurement tools and move the field of performance assessment in health care forward.

Editor's Note: A commentary by M.A. Rosen and P.J. Pronovost appears on pages [XXX–XXX].

Assessing clinical performance in health care is important for many reasons.[1] Doing so helps to characterize the abilities of clinicians and identify potential performance gaps. Further, assessing performance augments debriefings and forms the basis of scientific studies investigating factors influencing clinical performance. Because performance is a complex concept[2] and no single "best" performance measure exists, the reliable and valid assessment of performance is challenging for educators and researchers alike.

Please see the end of this article for information about the authors.

Correspondence should be addressed to Mr. Schmutz, University of Fribourg, Department of Psychology, Industrial Psychology and Human Factors, Rue de Faucigny 2, 1700 Fribourg, Switzerland; telephone: (+41) 26-300-7484; e-mail: jan.schmutz@unifr.ch.

Acad Med. 2014;89:00–00.
First published online
doi: 10.1097/ACM.0000000000000289

Systematic Performance Assessment

Within organizational psychology, performance is viewed as a multidimensional concept that comprises a process and an outcome component.[3,4] *Process performance* refers to what an individual or a team *does* in the work situation (e.g., performing a treatment task), whereas *outcome performance* refers to the *result* of this behavior (e.g., treatment-related patient outcomes).[2,4] Outcome performance measures, such as infection rate or mortality, can be assessed objectively but cannot always be directly attributed to clinical performance. For example, a patient might die despite a team's optimal performance in the resuscitation. Furthermore, in training settings, educators focus on correct clinical behaviors because outcomes are often not available and feedback on performance can modify trainees' behaviors. Thus, measures of process performance play a central role when assessing a trainee's clinical competence.

Process performance can be assessed by subjective and objective measures.[1] *Subjective measures*, which include global expert ratings of specific behavioral aspects or of overall performance, are mainly based on the clinical expertise of the rater. *Objective measures* are based on predefined scoring categories in the form of listed key items (i.e., evaluation checklists).[5]

This report focuses on checklists for evaluating clinical performance rather than on checklists supporting procedural task execution (e.g., central line placement). Checklists for evaluating clinical performance are structured tools outlining criteria to consider for a specific process.[6] They ensure that the assessment includes all important tasks during a process and, thus, force the rater to focus on predefined items. In the evaluation process, defining the specific criteria for the evaluation is crucial.[7] These criteria help to reduce observation biases (e.g., halo effect, confirmation bias),[8] and they can increase reliability among different evaluators.[9,10]

A classic evaluation checklist uses simple dichotomous items (done/not done). Because dichotomous items are frequently not sufficient for the assessment of complex tasks, this format has been extended to include more categories (e.g., done/done incorrectly/not done).[11–13] Other investigators have weighted checklist items to differentiate between essential

<zjs;Research Report> • <zsjs;XXX> • <z10.1097/ACM.0000000000000289>

LWW | 04/17/14 | 4 Color Fig(s):0 | 10:3 | Art: ACADMED-D-13-01053

and less important actions.[14] A few have defined specific actions as mandatory, which renders a total performance score of zero when the mandatory actions are not executed even if other actions are performed correctly.[15] One frequent criticism of evaluation checklists is that they reward thoroughness without considering the timeliness of actions.[16,17] Some researchers have acknowledged this by integrating time frames.[11,18] Factors such as weighting and time frames help create a more refined assessment of performance and should thus be considered in the development of future evaluation checklists.

Developing Checklists for Evaluating Clinical Performance

The development process of an evaluation checklist affects its quality.[19] The literature provides methodological recommendations for developing effective evaluation checklists: They should be based on (1) professional experience,[6,19] (2) primary literature sources or peer-reviewed guidelines,[19] and (3) the consensus of experts in the field of interest.[8,19] Table 1 provides examples of checklists, all of which have incorporated methodological recommendations from the literature and most of which also relied on expert opinion.

Not all studies in Table 1 include a description of a structured procedure for checklist development (i.e., defining a series of steps to be completed), nor do they all follow an overall systematic approach (i.e., defining guidelines or criteria for each of those steps). In fact, because the main focus of these studies is the evaluation of the checklist itself, only a few of them explicitly describe a structured and systematic approach to the checklist's development.[14,20,21] Particularly the later steps in the development process, such as weighting checklist items and integrating feedback from pilot testing, seem underemphasized (see Table 1).

Given the state of research and current practice in checklist development, we believe that researchers need a clear outline of the methodological steps to develop checklists for evaluating clinical performance, an outline that integrates existing recommendations into a more comprehensive approach. This integrated approach will support researchers in evaluating the suitability of checklists for different contexts, in designing performance assessment tools for specific clinical scenarios that reflect precisely what the task demands of the clinicians, and in either adapting existing checklists or generating new ones.

The aim of this study is to examine such an integrative approach in the development of checklists for evaluating clinical performance. Using the example of a simulated sepsis scenario, we have applied a five-step approach to checklist development that includes an adapted Delphi process and yields more than a classical dichotomous checklist by integrating timeliness and weights indicating the importance of different actions. In doing so we have used existing guidelines and methods and have integrated them into a comprehensive development process.

Method

This study was exempt from ethics review, per Swiss law. Figure 1 outlines the five steps of our systematic approach for the development of performance checklists. We developed and tested the five-step approach between May 2012 and June 2013.

Step 1: Development of a draft checklist

We developed an evaluation checklist for the simulated scenario of septic shock in a six-month-old boy. Three experienced acute care pediatricians and simulation educators (E.H., F.H., and W.J.E.) drafted an initial checklist of critical treatment tasks for this scenario based on published European Resuscitation Council guidelines,[22] the literature, and their own clinical experience.

Step 2: Delphi review rounds

We sent the draft checklist to five experts, whom we had chosen on the basis of established selection criteria, for review. The experts used an adapted Delphi technique to review the draft checklist. The Delphi technique is a consensus-based method through which experts respond to questionnaires and receive anonymous group feedback.[23] The main advantage of this method is the application of "collective intelligence," which is the combined ability of group members to jointly produce better results than anyone in the group could produce on his or her own.[24] The procedure consists of multiple review rounds until consensus is achieved. When Delphi rounds are conducted by mail or e-mail, the process is anonymous, allowing each expert to make suggestions without fear of losing face. The anonymity also reduces the impact of common group biases like conformity or power influences.[23] The Delphi technique is well established in social science and increasingly used in health care research for various purposes.[25–27]

Selection of experts. Recommended sample sizes for experts for a Delphi study range from 5 to 30 depending on the research question.[28] In line with these recommendations, we felt a sample of 5 experts would be sufficient because the treatment of septic shock mostly follows established, standardized algorithms. The validity of the Delphi technique depends strongly on the selection of the experts; thus, we required all experts to be board-certified physicians with at least 10 years of clinical practice after medical school including at least 6 years in pediatric care.

Procedure. Figure 1 provides a visual representation of the Delphi review rounds. Experts received the draft checklist as well as a short history of the simulated patient and the current sepsis scenario (Box 1) by e-mail. We instructed the experts not only to delete irrelevant actions, add missing but relevant actions, or reformulate already-listed items but also to include a comment explaining all additions, deletions, and reformulations as information for all experts in the next review round. After the first round, all expert feedback was integrated into one modified list, and the source of all edits was deidentified. All suggestions were clearly highlighted.

In round 2, we asked participants to confirm whether or not an added item should remain in the list and if they agreed with the proposed deletions. If a majority (three of five experts) recommended an addition or deletion, we included the change. We repeated this procedure until the experts achieved a consensus and made no more suggestions.

Step 3: Design of the final checklist and pilot testing

In the third step, three clinicians, including two of us (E.H. and F.H.),

Table 1

Characteristics of Procedures Used for Designing Performance Checklists in the Literature

Study	Scenario evaluated by checklist	Literature- and/or guideline based	Expert opinion based	Structured development process[a]	Overall systematic approach[b]
Chopra et al, 1994[43]	—Anaphylactic shock —Malignant hyperthermia	✓			
Gaba et al, 1998[15]	—Malignant hyperthermia —Cardiac arrest	✓	✓		
Lockyer et al, 2006[20]	—Neonatal resuscitation	✓	✓	✓	✓
Scavone et al, 2006[21]	—General anesthesia for emergency cesarean delivery	✓	✓ (Delphi)	✓	✓
Thomas et al, 2006[44]	—Neonatal resuscitation	✓			
Tschan et al, 2006[45]	—Cardiac arrest	✓			
Adler et al, 2007[46]	—Apnea —Asthma —Supraventricular tachycardia —Sepsis	✓	✓	✓ (CDC)[19]	
Morgan et al, 2007[14]	Anesthesia induction for patient with: —Laparoscopic cholecystectomy —Laparotomy for large bowel obstruction	✓	✓ (Delphi)	✓	✓
Brett-Fleegler et al, 2008[47]	—Near-drowning child —Child with asthma —Child with tricyclic antidepressant overdose	✓	✓ (Delphi and development session)	✓	
Adler et al, 2009[48]	—Infant in shock —Tachycardia —Altered mental status —Trauma	✓	✓	✓ (CDC)[19]	
Carlson et al, 2009[49]	—Acute dyspnea	✓			
Donoghue et al, 2009[11]	—Asystole —Tachydysrhythmia —Respiratory arrest —Shock	✓	✓	✓	✓ (Approach by Lockyer)[20]
Manser et al, 2009[50]	—Anesthesia induction in malignant hyperthermia	✓			
Burtscher et al, 2010[51]	—Standard anesthesia induction	✓	✓ (Delphi)	✓	
Donoghue et al, 2010[18]	—Asystole —Dysrhythmia —Respiratory arrest —Shock	✓	✓	✓	✓ (Approach by Lockyer)[20]
Westli et al, 2010[52]	—Trauma	✓			
Adler et al, 2011[16]	—Shock —Unexplained altered mental status —Multisystem trauma	✓	✓	✓ (CDC)[19]	
Burtscher et al, 2011[40]	—General anesthesia induction	✓	✓ (Delphi)	✓	
Lambden et al, 2013[32]	—Respiratory failure —Sepsis —Meningitis with raised intracranial pressure	✓	✓		

Abbreviations: CDC indicates Checklists Development Checklist.
[a]Following a well-predefined process.
[b]Methodical approach predefined and replicable through a step-by-step procedure.

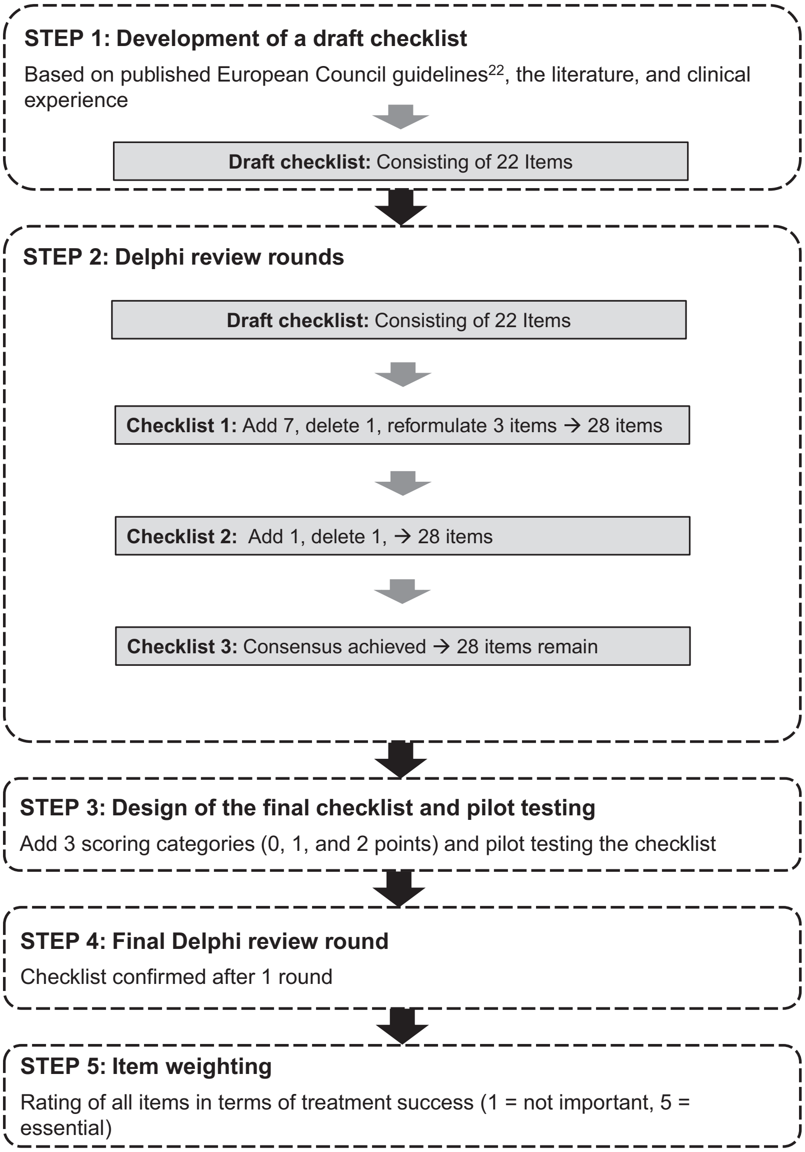

Figure 1 The five steps to develop checklists for evaluating clinical performance.

pilot tested the checklist by rating the videotaped management of six simulated pediatric septic shock scenarios. Through this process, we identified items that were formulated ambiguously, items that could not clearly be observed (e.g., items referring to cognitive processes), and problems in the order or grouping of items.

To increase the accuracy of the evaluation of a performance, we followed the example of the Clinical Performance Tool[29] and specified, for each checklist item, three scoring categories: task not performed (zero points); task performed partially, incorrectly, or with delay (one point); and task performed completely, correctly, and within the recommended time frame (two points). For example, a team would score two points for calling for help in the first five minutes but only one point for doing so after five minutes.

Box 1

Patient History of the Simulated Septic Shock Scenario

A 6-month-old male infant presents with several hours of fever and vomiting. The fever responds poorly to antipyretics, and the infant becomes progressively lethargic and responds only to painful stimulus. Skin exam reveals scattered nonblanching petechiae. Two minutes after initial evaluation, the infant becomes unresponsive.

Step 4: Final Delphi review round

To ensure expert consensus concerning changes made in step 3, we sent the revised checklist for review, asking the original five pediatrician experts, as before, to delete irrelevant actions, add missing items, and/or edit listed items.

Step 5: Item weighting

Not every item in a checklist is equally important for the treatment to be successful. A checklist differentiating between essential and less important items is likely to provide more accurate performance assessments. Thus, in a final development step, we sent the checklist to 10 pediatricians and pediatric anesthetists from Switzerland, Germany, and Australia. We instructed them to rate all checklist items in terms of their importance for the success of the treatment from 1 (not important) to 5 (essential). The mean importance score of each item served as its weight.

Internal consistency and validity

We tested internal consistency of the final checklist by having three pediatricians, including two of us (E.H. and F.H.) and an independent rater, assess four videotaped samples of managing the simulated pediatric septic shock scenario. We assessed Cohen kappa[30] to measure agreement (for all four videos) between the independent rater and either E.H. or F.H.

We assessed *content* validity (the extent to which the checklist includes all relevant items) through a detailed discussion at an international workshop for simulation in medicine.

Evidence for *construct* validity would mean that the checklist score positively but not excessively correlated with the external constructs.[31] We applied two external constructs commonly used for validation[12,32]: (1) a global expert performance rating from 1 to 10 (given by E.H., F.H., and the independent rater before rating the video with the evaluation checklist) and (2) the experience level of the team leader (assessed by a questionnaire; in an emergency scenario, a team with a more experienced leader should get higher scores than a team with an inexperienced leader). We tested construct validity using a sample of 22 teams performing a septic shock scenario in a simulated setting.

Results

Step 1: Development of a draft checklist

Three of us (E.H., F.H., and W.J.E.) developed a draft checklist consisting of 22 potential items.

Step 2: Delphi review rounds

Experts. The five experts included in the Delphi process had 14 to 28 years of general medical experience and had worked 6 to 27 years in pediatric care in different Swiss and German hospitals.

Delphi rounds and checklist changes. During the first review round, experts made the following suggestions: seven items for addition; one item for deletion; and three items for reformulation.

In the second round, all the experts agreed to add six of the seven items newly suggested for addition in round 1. Four of the five experts did not agree with the seventh addition, so this one item was excluded. The majority (n = 4) of experts disagreed with the proposal to exclude the one item suggested for deletion in round 1; thus, this item was included again. All the experts agreed with the proposed rewording of three items. Furthermore, one new additional item was proposed to add to the list.

After the third round, all the experts agreed to add the additional item suggested in round 2 and had no further suggestions. Also, the two changes which were not accepted by the majority in round 2 (deleting one item and adding one item) were, at this point, accepted by the corresponding expert who had proposed the changes based on the detailed comments of opposing experts. So the five experts achieved consensus about all the items in the list after three review rounds. The list, after step 2, contained 28 items.

Step 3: Design of the final checklist and pilot testing

At this step, we determined which of the 28 items made sense to rate with the three scoring categories. For 7 of the items, the scoring option "partially done" made no sense (e.g., check pulse, check temperature), leaving the option of only zero or two points.

Using the checklist, two of us (E.H. and F.H.) individually rated the video-recorded management of six simulated cases of septic shock and took notes about problems with specific scoring categories. Next, we discussed possible improvements to the checklist. We identified four types of adjustments to enhance the usability of the checklist. Table 2 provides the four types of adjustments made and the corresponding scoring categories. After the pilot phase, the checklist contained 33 items. T2

Step 4: Final Delphi review round

All five experts agreed with all adjustments made in step 3.

Step 5: Item weighting

The average experience after medical school of the 10 experts participating in step 5 was 16 years (standard deviation [SD] = 9.9), and in a pediatric field specifically, it was 11.4 years (SD = 8).

Internal consistency and validity

The mean score of the ratings ranged from 3 to 5. In general, the SD was small. Only 6 of the 33 items had an SD of more than 1.0. The two items least specifically related to the immediate treatment of septic shock generated the highest disagreement: "Put on gloves before procedure" (SD = 1.73) and "Early planning for other treatment" (SD = 1.35). The final list including the rounded weighting of each item can be seen in Appendix 1.

Interrater reliability analyses of the resulting checklist revealed "substantial" to "almost perfect" kappa coefficients[33] (κ range: 0.65–0.95).

Our thorough, integrative development process through which we derived the items provides content validity.[1] Further, participants of an international workshop for simulation in medicine agreed that the content of the checklist includes all necessary items. The correlation between the checklist score and team leader experience ($r = 0.48$, $P < .05$) and the global rating ($r = 0.68$, $P < .05$) were both significant. Thus, the checklist yields valid results.

Discussion

To design effective training interventions, valid and reliable performance measures must be developed systematically. In this report, we describe a systematic approach to designing checklists for evaluating clinical performance that integrates the published evidence and the knowledge of domain experts. Through its clearly predefined procedure, our method reduces opportunities for subjective interpretations and thus minimizes rater biases. Our approach consists of five easy-to-apply, predefined steps that integrate the following: current checklist development guidelines, an expert consensus method, pilot testing, an additional expert consensus round, and a

Table 2

The Four Types of Adjustments Made After Pilot Testing the Checklist for Taking Care of an Infant With Septic Shock, 2013

Type of adjustment	Problem	Old item	Solution or new item(s)
Specifying	No clear and objective definition of item	"Connect monitors"	Connect ECG, SpO_2, and blood pressure monitors
Splitting	Items are formulated too broadly	"Order and give fluid bolus 3 times"	—Order first fluid bolus —Give first fluid bolus —Order second fluid bolus —Give second fluid bolus —Order third fluid bolus —Give third fluid bolus
Eliminating redundancy	Different items include the same actions	"ABC evaluation"	Item deleted because the following items were already in the list: —Assess airway/breathing —Assess mental status
Changing the order of items	Inconvenient order of items due to thematic grouping instead of grouping according to the course of events		The order should correspond to the expected course of events as much as possible so as to minimize rater search time

Abbreviations: ECG indicates electrocardiogram; SpO_2, oxygen saturation; ABC, airway, breathing, circulation.

survey to get importance ratings for the checklist items.

Our approach has some advantages over other systematic approaches.[14,20,21] Lockyer and colleagues[20] proposed a method to develop a checklist in three stages. In stage 1, the authors created an evaluation checklist and then published it on the Web site of the Neonatal Resuscitation Program (NRP) for additional review. Then the NRP recruited volunteers to review the list by mail. In stage 3, a pilot test was conducted in which experienced instructors used the list to rate specific video clips. After each step the checklist was modified. Although Lockyer et al[20] obtained feedback from a large number of responders and conducted a pilot test, it is unclear how they modified the list after every step and how they dealt with conflicting comments. Further consensus methods were not applied.

Morgan et al[14] and Scavone et al[21] both used a well-defined Delphi technique. They required experts participating in their Delphi technique to agree not only with the items included in the checklist but also to a weight (of 1 to 5) for each item. These weights could be problematic because the two analyses did not consider the small sample size and because the final checklist used the mean score of the expert ratings for the items for which no consensus could be achieved. Therefore, we strongly suggest conducting a separate step (following our step 5) to obtain the weights of the checklist items, allowing for an adequate sample size, at least for those items for which no consensus is achieved.

All three aforementioned studies included a pilot phase through which raters tested the checklist as an assessment tool.[14,20,21] This step is indispensable; by applying the checklist to a set of different examples, the raters experience the applicability of the items and the usability of the rating scale. Each item has to be formulated in a clear and observable way. If it is not, then it must be excluded or modified so that it does not threaten interrater agreement. For example, the item "Equipment check" seems absolutely reasonable. However, to get reliable ratings, the checklist must define *what* equipment has to be checked (e.g., oxygen connector, ventilation bag). Another problem arises with the rating of behaviors that are hard to detect or are executed mentally (e.g., "Check breathing"—whether or not the trainee has perceived the lifting and lowering of the chest is unclear if he or she does not verbalize doing so).

After the pilot phase, we conducted a final Delphi review round (step 4). To our knowledge, no other study has included additional expert feedback *after* a pilot phase. This step is important because it ensures that the adjustments made after the first testing are recognized by experts and not biased in any way by raters' personal opinions or by experiences that are not generally valid.

Lessons learned

Not only the Delphi review rounds but also the inquiry about the item weights can take a long period of time. Content experts in the field are often very busy clinicians, and responding to the inquiries is not their first priority. If possible, checklist developers and investigators should consider creating individual incentives for the experts to enhance their commitment (e.g., free access to the final product).

Further, we noted some process issues: In one case, we detected a lack of expert diligence in providing feedback, and in another an expert overlooked some items and did not comment on them. Soliciting the missing comments lengthened the time of that particular Delphi review round. Thus, we recommend emphasizing the importance of the experts' contribution in the first communication and indicating a reasonable expectation for response time so that experts can reject the invitation immediately rather than dropping out later.

Areas of application

We demonstrated our approach using the example of a simulated sepsis scenario. Our approach, though, is not limited to one scenario; it is generalizable. We have since successfully applied this five-step process to other clinical scenarios (i.e., pulseless ventricular tachycardia, bronchiolitis, and near-drowning). In doing so, we have created checklists which correspond to the specific context in question and which, in some cases, differ considerably from the initial checklist drafted by the research team.

Evaluation checklists are generally most suitable for training purposes or for simulated scenarios in which no patient outcome measures are available. Our approach to checklist development may be particularly useful to design evaluation checklists for situations that have a certain degree of standardization and are frequently covered by guidelines. Because there are national differences in the treatment of specific clinical scenarios, our approach can also be employed to include expert feedback when adapting existing checklists for a new national setting. Our approach would also be useful in any setting for updating a list to account for changes in guideline regulations. For less standardized situations, in which the actions depend highly on the particular situation, assessing performance with a checklist may not be suitable; for example, a checklist would not capture the many skills necessary for managing a critically ill child with a complex past medical history and dealing with end-of-life issues related to "do not resuscitate" orders or withdrawal of intensive care. In such situations, another form of assessment, such as a behaviorally anchored rating scale, a global rating tool, or patient-focused outcomes, may augment performance assessment.

Physicians and physician educators can use our five-step procedure not only to design checklists for performance assessment but also for the development of cognitive aids that help ensure all necessary tasks are completed.[34] The use of checklists has been demonstrated to reduce error by standardizing specific processes in surgery,[35] anesthesia,[36] handover,[37] and inpatient care.[38]

Limitations

Despite the advantages of our approach, we note some limitations. The development of an evaluation checklist according to our approach requires significant time and effort. Patient outcomes, specific performance markers (e.g., time to key interventions,[39] decision latency),[40] and global rating scales are often easier to assess and do not require a long development process. Thus, some researchers propose global rating scales as the preferred assessment tool.[16,41] However, patient outcomes or global ratings often do not provide comprehensive evaluations of the treatment process and, thus, cannot provide process feedback to augment debriefings.

One notable limitation concerns the fact that two raters involved in pilot testing (E.H., F.H.) were both also involved in the development process. The testing of a new tool should ideally be done by independent potential users.[42] In our case, we were able to show good interrater agreement with a third, independent rater during the validation process. Therefore, we believe that this limitation had no negative effect on the final checklist. Nevertheless, we recommend independent raters during pilot testing.

Other limitations are related to the Delphi technique in general. Although this process facilitates reaching expert consensus, it does not necessarily mean that this consensus is "correct." Although the possibility of the consensus being influenced by a single expert's opinion is small, the possibility still cannot be ruled out completely. In our case, there was no serious disagreement about whether an individual item should be included in the checklist or not; thus, we assume that this issue did not influence our results.

Further, the country of origin and professional background of the experts could influence their responses. Medical guidelines may vary on a regional or national basis; even local factors at an individual hospital can result in different expert opinions. Although completely controlling for the background of every expert is almost impossible, we tried to counteract cultural differences by selecting the experts from regions where the final checklist should be applied (Germany, Switzerland, Australia). Differences in culture and regions should be kept in mind when choosing the experts for future studies.

Finally, we developed the evaluation checklist for a specific pediatric sepsis scenario as we use it in our simulation trainings. This local context might have influenced the development process in a way that may preclude adopting the final checklist for other sepsis scenarios without making small adjustments.

Future research

In future studies, other formats and venues for performing the Delphi review rounds and their impact on the quality of the final checklist should be explored. A consensus meeting instead of e-mail inquiry may result in a more dynamic and deliberate discussion of the checklist and would speed up the process. Clearly, a disadvantage of such a consensus meeting could be the higher risk of group biases because the experts would no longer be anonymous. To avoid this potential drawback, a consensus meeting could be held online in which experts could maintain anonymity in a virtual chat room.

Conclusions

Assessment of clinical performance is fundamental to further enhance patient safety. Only reliable and valid process performance measures that are less influenced by unknown variables (than are clinical outcomes) will allow medical educators to accurately evaluate the behavioral effects of training interventions and, in turn, leverage and modify the training.

A systematic development process is a necessary prerequisite of valid checklists for reliably assessing process performance. However, no universally agreed guideline for the systematic development of evaluation checklists exists. With this report, we describe a comprehensive integrative approach that may be used in future studies. We are convinced that a widely recognized standard for developing evaluation checklists, such as the one we applied, would advance the field of performance assessment in health care.

Acknowledgments: The authors are grateful to all experts for their participation in the development process of the checklist. They also thank Lauren Clack for her help with the design of the figures and Julia Keil for her participation as an independent rater for the reliability analysis.

Funding/Support: This work was funded by the Swiss National Science Foundation (grant number PP00P1_128616).

Other disclosures: None.

Ethical approval: This type of research is exempt from ethics review in Switzerland.

Disclaimer: None.

Previous presentations: Oral presentation at the 19th Annual Meeting of the Society in Europe for Simulation Applied to Medicine (SESAM), June 13 to 15, 2013, Paris, France.

Mr. Schmutz is research associate, Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Fribourg, Switzerland.

Dr. Eppich is assistant professor of pediatrics and medical education, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

Dr. Hoffmann is senior pediatrician, Dr. von Hauner University Children's Hospital, Munich, Germany.

Dr. Heimberg is pediatrician, University Children's Hospital, Tübingen, Germany.

Dr. Manser was, at the time of this research, associate professor for industrial psychology and human factors, Department of Psychology, University of Fribourg, Fribourg, Switzerland, and is currently professor of patient safety, Faculty of Medicine, University of Bonn, Bonn, Germany.

References

1 Boulet JR, Murray D. Review article: Assessment in anesthesiology education. Can J Anaesth. 2012;59:182–192.
2 Campbell. Modeling the performance prediction problem in industrial and organizational psychology. In: Dunnette MD, Hough LM, eds. Handbook of Industrial and Organizational Psychology. Palo Alto, Calif: Consulting Psychologists Press; 1990:687–732. AQ1
3 Sonnentag S, Frese M. Performance concepts and performance theory. In: Sonnentag S, ed. Psychological Management of Individual Performance. West Sussex, UK: John Wiley & Sons, Ltd.; 2002:1–25.
4 Anderson N, Ones DS, Sinangil HK, Viswesvaran C. Handbook of Industrial, Work and Organizational Psychology: Personnel Psychology. Vol 1. London, UK: Sage Publications Ltd.; 2001.
5 Boulet JR, Jeffries PR, Hatala RA, Korndorffer JR Jr, Feinstein DM, Roche JP. Research regarding methods of assessing learning outcomes. Simul Healthc. 2011;6(suppl):S48–S51.
6 Hales B, Terblanche M, Fowler R, Sibbald W. Development of medical checklists for improved quality of patient care. Int J Qual Health Care. 2008;20:22–30.
7 Davidson JE. Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation. Thousand Oaks, Calif: Sage; 2004.
8 Scriven M. The logic and methodology of checklists. December 2007. http://www.wmich.edu/evalctr/archive_checklists/papers/logic&methodology_dec07.pdf. Accessed March 25, 2014.
9 Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. Acad Med. 2001;76:1053–1055.
10 Bakeman R, Gottman JM. Observing Interaction: An Introduction to Sequential Analysis. 2nd ed. Cambridge, UK: Cambridge University Press; 1997.
11 Donoghue JD, Durbin DR, Nadel FM, Stryjewski GR, Kost SI, Nadkarni VM. Effect of high-fidelity simulation on pediatric advanced life support training in pediatric house staff. Pediatr Emerg Care. 2009;25:139–144.
12 Devitt JH, Kurrek MM, Cohen MM, et al. Testing internal consistency and construct validity during evaluation of performance in a patient simulator. Anesth Analg. 1998;86:1160–1164.

13 Reid J, Stone K, Brown J, et al. The Simulation Team Assessment Tool (STAT): Development, reliability and validation. Resuscitation. 2012;83:879–886.
14 Morgan PJ, Lam-McCulloch J, Herold-McIlroy J, Tarshis J. Simulation performance checklist generation using the Delphi technique. Can J Anaesth. 2007;54:992–997.
15 Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. Anesthesiology. 1998;89:8–18.
16 Adler MD, Vozenilek JA, Trainor JL, et al. Comparison of checklist and anchored global rating instruments for performance rating of simulated pediatric emergencies. Simul Healthc. 2011;6:18–24.
17 Cohen R, Rothman AI, Poldre P, Ross J. Validity and generalizability of global ratings in an objective structured clinical examination. Acad Med. 1991;66:545–548.
18 Donoghue A, Nishisaki A, Sutton R, Hales R, Boulet J. Reliability and validity of a scoring instrument for clinical performance during pediatric advanced life support simulation scenarios. Resuscitation. 2010;81:331–336.
19 Stufflebeam DL. Guidelines for developing evaluation checklists: The checklists development checklist (CDC). July 2000. http://www.wmich.edu/evalctr/archive_checklists/guidelines_cdc.pdf. Accessed March 25, 2014.
20 Lockyer J, Singhal N, Fidler H, Weiner G, Aziz K, Curran V. The development and testing of a performance checklist to assess neonatal resuscitation megacode skill. Pediatrics. 2006;118:e1739–e1744.
21 Scavone BM, Sproviero MT, McCarthy RJ, et al. Development of an objective scoring system for measurement of resident performance on the human patient simulator. Anesthesiology. 2006;105: 260–266.
22 Biarent D, Bingham R, Eich C, et al. European Resuscitation Council Guidelines for Resuscitation 2010 Section 6. Paediatric life support. Resuscitation. 2010;81:1364–1388.
23 Gordon TJ. The Delphi Method. American Council for the United Nations University (AC/UNU) Millenium Project. 1994. http://fpf.ueh.edu.vn/imgnews/04-Delphi.pdf. Accessed March 25, 2014.
24 Turoff M, Hiltz SR. Computer based Delphi processes. In: Adler M, Ziglio E, eds. Gazing Into the oracle: The Delphi Method and Its Application to Social Policy and Public Health. London, UK: Kingsley Publishers Ltd.; 1996:56–88.
25 Ferguson ND, Davis AM, Slutsky AS, Stewart TE. Development of a clinical definition for acute respiratory distress syndrome using the Delphi technique. J Crit Care. 2005;20:147–154.
26 Huang HC, Lin WC, Lin JD. Development of a fall-risk checklist using the Delphi technique. J Clin Nurs. 2008;17:2275–2283.
27 Naik VN, Perlas A, Chandra DB, Chung DY, Chan VW. An assessment tool for brachial plexus regional anesthesia performance: Establishing construct validity and reliability. Reg Anesth Pain Med. 2007;32:41–45.
28 Clayton MJ. Delphi: A technique to harness expert opinion for critical decision-making tasks in education. Educ Psychol. 1997;17:373–386.
29 Donoghue A, Ventre K, Boulet J, et al; EXPRESS Pediatric Simulation Research Investigators. Design, implementation, and psychometric analysis of a scoring instrument for simulated pediatric resuscitation: A report from the EXPRESS pediatric investigators. Simul Healthc. 2011;6:71–77.
30 Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
31 Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52:281–302.
32 Lambden S, DeMunter C, Dowson A, Cooper M, Gautama S, Sevdalis N. The Imperial Paediatric Emergency Training Toolkit (IPETT) for use in paediatric emergency training: Development and evaluation of feasibility and validity. Resuscitation. 2013;84:831–836.
33 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174.
34 Hales BM, Pronovost PJ. The checklist——a tool for error management and performance improvement. J Crit Care. 2006;21:231–235.
35 Haynes AB, Weiser TG, Berry WR, et al; Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. N Engl J Med. 2009;360:491–499.
36 Myburgh JA, Chapman MJ, Szekely SM, Osborne GA. Crisis management during anaesthesia: Sepsis. Qual Saf Health Care. 2005;14:e22.
37 Foster S, Manser T. The effects of patient handoff characteristics on subsequent care: A systematic review and areas for future research. Acad Med. 2012;87:1105–1124.
38 Wolff AM, Taylor SA, McCabe JF. Using checklists and reminders in clinical pathways to improve hospital inpatient care. Med J Aust. 2004;181:428–431.
39 Marsch SC, Müller C, Marquardt K, Conrad G, Tschan F, Hunziker PR. Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. Resuscitation. 2004;60:51–56.
40 Burtscher MJ, Manser T, Kolbe M, et al. Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. Br J Anaesth. 2011;106:1–6.
41 Ma IW, Zalunardo N, Pachev G, et al. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. Adv Health Sci Educ Theory Pract. 2012;17:457–470.
42 Bevan N, Macleod M. Usability measurement in context. Behav Inf Technol. 1994;13:132–145.

References That Appear in Tables Only

43 Chopra V, Gesink BJ, de Jong J, Bovill JG, Spierdijk J, Brand R. Does training on an anaesthesia simulator lead to improvement in performance? Br J Anaesth. 1994;73:293–297.
44 Thomas EJ, Sexton JB, Lasky RE, Helmreich RL, Crandell DS, Tyson J. Teamwork and quality during neonatal care in the delivery room. J Perinatol. 2006;26:163–169.
45 Tschan F, Semmer NK, Gautschi D, Hunziker P, Spychiger M, Marsch SU. Leading to recovery: Group performance and coordinative activities in medical emergency driven groups. Hum Per. 2006;19:277–304.
46 Adler MD, Trainor JL, Siddall VJ, McGaghie WC. Development and evaluation of high-fidelity simulation case scenarios for pediatric resident education. Ambul Pediatr. 2007;7:182–186.
47 Brett-Fleegler MB, Vinci RJ, Weiner DL, Harris SK, Shih MC, Kleinman ME. A simulator-based tool that assesses pediatric resident resuscitation competency. Pediatrics. 2008;121:e597–e603.
48 Adler MD, Vozenilek JA, Trainor JL, et al. Development and evaluation of a simulation-based pediatric emergency medicine curriculum. Acad Med. 2009;84:935–941.
49 Carlson J, Min E, Bridges D. The impact of leadership and team behavior on standard of care delivered during human patient simulation: A pilot study for undergraduate medical students. Teach Learn Med. 2009;21:24–32.
50 Manser T, Howard SK, Gaba DM. Identifying characteristics of effective teamwork in complex medical work environments: Adaptive crew coordination in anaesthesia. In: Flin R, Michell L, eds. Safer Surgery: Analysing Behaviour in the Operating Theatre. Aldershot, UK: Ashgate; 2009:223–239.
51 Burtscher MJ, Wacker J, Grote G, Manser T. Managing nonroutine events in anesthesia: The role of adaptive coordination. Hum Factors. 2010;52:282–294.
52 Westli HK, Johnsen BH, Eid J, Rasten I, Brattebø G. Teamwork skills, shared mental models, and performance in simulated trauma teams: An independent group design. Scand J Trauma Resusc Emerg Med. 2010;18:47.

Appendix 1

Checklist Developed[a] to Evaluate Care of an Infant With Septic Shock, 2012

Stage of care (timing in minutes)	Item no.	Item description	Scoring (Check the box for 0, 1, or 2 points, as appropriate)						Weighting
			Not done (0 points)	Partially or incorrectly done or not done in a timely manner (1 point)		Done correctly, completely, and in a timely manner (2 points)			
General tasks (0–5)	1-1	Put on gloves before procedure	☐	Some, but not all persons involved in procedure put on gloves	☐	Every person who is involved in procedure puts on gloves	☐		3
	1-2	Equipment check	☐	Incomplete: Oxygen connected *or* ventilator bag checked	☐	Oxygen connected, ventilation bag checked	☐		3.5
	1-3	Connect ECG, SpO_2, BP	☐	Incomplete (only 1 or 2 items)	☐	All complete	☐		4.5
	1-4	Call for help (senior physician)	☐	Not in time (i.e., after actors' recommendation)	☐	Done in time	☐		5
	1-5	Inform team members about diagnosis	☐	Not done in time *or* incomplete information related to vital signs (i.e., only "tachycardia" or only "low blood pressure")	☐	Complete information about diagnosis "shock"	☐		4
Evaluation (0–5)	2-1	Assess airway and breathing	☐	Only bilateral auscultation ***or*** assess breathing frequency ***or*** work of breathing	☐	Bilateral auscultation ***and*** assess breathing frequency ***and*** work of breathing	☐		5
	2-2	Check SpO_2	☐	Done	☐	Done in time and verbalized	☐		3.5
	2-3	Check pulse	☐			Done in time	☐		4
	2-4	Check ECG	☐	Done	☐	Done in time and verbalized	☐		3.5
	2-5	Check CRT	☐			Done in time	☐		4.5
	2-6	Check BP	☐	Done	☐	Done in time and verbalized	☐		4
	2-7	Check temperature	☐			Done in time	☐		3
	2-8	Assess mental status	☐			Done (explicit question about mental status, e.g., AVPU)	☐		4.5
Treatment (0–5)	3-1	Apply oxygen	☐	Nasal cannula	☐	100% O_2 applied	☐		4.5
	3-2	Establish IV/IO access	☐	More than 2 attempts for IV access ***or*** not in time	☐	Successful in maximum of two IV attempts	☐		5
	3-3	Order first fluid bolus	☐	Wrong fluid ***or*** wrong amount ordered *or* not in time	☐	Right dose (20 mL/kg) and right fluid	☐		5
	3-4	Start giving first fluid bolus	☐	IV pump *or* rapid IV push not in time	☐	Rapid IV push	☐		5

(*Appendix Continues*)

Appendix 1

(Continued)

Stage of care (timing in minutes)	Item no.	Item description	Scoring (Check the box for 0, 1, or 2 points, as appropriate) Not done (0 points)	Partially or incorrectly done or not done in a timely manner (1 point)	Done correctly, completely, and in a timely manner (2 points)	Weighting
Treatment and assessment (5–15)	4-1	Reassess circulation (CRT, BP, HR)	☐	Incomplete (checked only 1 or 2) ☐	All complete ☐	4.5
	4-2	Reassess breathing (SpO_2, breathing frequency)	☐	Incomplete 1 (checked only 1) ☐	All complete ☐	5
	4-3	Order second fluid bolus	☐	Wrong fluid *or* wrong amount ordered ☐	Right dose and right fluid ☐	4.5
	4-4	Give second fluid bolus	☐	IV pump ☐	Rapid IV push ☐	4.5
	4-5	Reassess circulation (CRT, BP, HR)	☐	Incomplete ☐	All complete ☐	5
	4-6	Reassess breathing (SpO_2, breathing frequency)	☐	Incomplete ☐	All complete ☐	5
	4-7	Order third fluid bolus	☐	Wrong fluid *or* wrong amount ordered ☐	Right dose and right fluid ☐	4.5
	4-8	Give third fluid bolus	☐	IV pump ☐	Rapid IV push ☐	4.5
	4-9	Reassess circulation (CRT, BP, HR)	☐	Incomplete ☐	All complete ☐	4.5
	4-10	Reassess breathing (SpO_2, breathing frequency)	☐	Incomplete	All complete ☐	4.5
Reassessment and planning for other treatment (5–15)	5-1	Reassess mental status	☐		All complete ☐	4
	5-2	Consider vasoactive agents	☐		Done ☐	4
	5-3	Consider antibiotics	☐	Administration discussed ☐	Ceftriaxon or Cefotaxim ordered ☐	5
	5-4	Draw blood culture and BGA glucose	☐	Incomplete ☐	Includes at least BGA, blood culture, electrolytes ☐	4
	5-5	Prepare for advanced airway management	☐	Incomplete ☐	Suction, medication, bag mask, laryngoscope ☐	4.5
	5-6	Early planning for other treatment	☐		Done ☐	3.5

Abbreviations: ECG indicates electrocardiogram; SpO_2, oxygen saturation; CRT, capillary refill time; BP, blood pressure; IV, intravenous; IO, intraosseous; HR, heart rate; AVPU, alert, responds to voice, responds to pain, unresponsive; BGA, blood gas analysis.

[a]The authors used an integrated, five-step approach to develop the checklist.

5.3 Study C: Coordination in healthcare action teams: Utilizing expert under-standing of task and team performance requirements

Sarah Henrickson Parker[1], Jan Schmutz[2], Tanja Manser[2]

[1]National Center for Human Factors in Healthcare, MedStar Health
[2]Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Switzerland

This manuscript was submitted for publication and is under revision at the Journal of Occupational and Organizational Psychology

INTRODUCTION

Teams are ubiquitous in high-risk, high-reliability organizations. Specifically within healthcare, the process of providing care is inherently interdisciplinary. Doctors, nurses, allied health professionals; specialists must all work together to achieve the common goal of caring for patients. It has been widely documented that in order to achieve safe patient care, effective teamwork is critical (Kohn, Corrigan, & Donaldson, 2000; Mishra, Catchpole, Dale, & McCulloch, 2008; Schmutz & Manser, 2013; Manser, 2009). Early work on understanding error in healthcare has established that many adverse events are related to communication and team performance, rather than to negligence or inadequate technical skills of caregivers (Helmreich, 2000). The ability to perform well within a complex organization is not exclusively dependent on individuals, but rather on their skills of working together as a team (Burke, Salas, Wilson-Donnelly, & Priest, 2004; Klein, Ziegert, Knight, & Xiao, 2006; Klinger & Thorsden, 1998; Mishra et al., 2008). There is considerable interest among caregivers, researchers, as well as among hospitals, insurance organizations, and most importantly from patients into how to optimize team performance and thus, decrease preventable harm.

In the psychology literature, teamwork has been described as the process of two or more individuals who work together to achieve task goals, have task-specific competencies and specialized work roles, use shared resources and communication to coordinate and adapt to change (Brannick & Prince, 1997). While this definition adequately describes most teams in most organizations, teams that function within emergency care settings, whose work is poorly defined compared to normal organizational standards, require additional description. Teams that work under changing conditions, may be assembled ad hoc, have dynamic membership, often work together for a short period of time, and involve specialists or specialist sub groups are considered "action teams" (Sundstrom, De Meuse, & Futrell, 1990).

Action teams have often been investigated within military contexts (Burke et al., 2004; Lim & Klein, 2006; Marks & Panzer, 2004; Naikar, Pearce, Drumm, & Sanderson, 2003) as well as in fire-fighting (Bigley & Roberts, 2001). These studies show that teams in these flexible, dynamic environments may have different team processes and structures that contribute to optimal performance than do typical teams. Within various healthcare settings, action teams are frequently established in to address immediate patient needs during a single procedure. Action teams have been studied in trauma resuscitation (Xiao, Seagull, Mackenzie, Klein, & Ziegert, 2002), cardiac resuscitation (Hunziker, Tschan, Semmer, & Marsch, 2013; Tschan et al., 2011a), anesthetic care (Burtscher, Zurich, & Wacker, 2010; Kolbe, Burtscher, Manser, & Barbara, 2011; Manser, Howard, & Gaba, 2008; Manser, Harrison, Gaba, & Howard, 2009), and surgery (Edmonson, 2003; Lingard et al., 2004; Wiegmann, Eggman, Elbardissi, Parker, & Sundt 3rd, 2010) among other care environments.

Although protocols for performance may exist, the task of action teams is evolving, thus requiring teams to engage in a "process by which team resources, activities, and responses are organized to ensure that tasks are integrated, synchronized, and completed within established temporal constraints" (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995, p. 345). These behaviours are often summarized under the term as "coordination" that is seen as a key team process. Coordination is the way in which teams organize or manage their actions and share information in order to achieve task goals (Kolbe et al., 2011; Tschan et al., 2011b). Although coordination is important for all teams, excellent coordination may be more important to effective performance under action team conditions.

Coordination is essential for expert team performance

Of critical importance for understanding and optimizing performance in action teams is the team's ability to coordinate their work effectively and adaptively. Action teams must collectively understand the task and its changing demands, continuously update and reassess information, and integrate new information into both task and team goal assessment. In a fast-paced environment, the team may have a clear goal, the right mix of expertise, appropriate resources and teamwork structures that support effectiveness, yet still suffer breakdowns in coordination due to miscommunication, interpersonal issues or poor judgement under stress (Weick, 1993). In action teams, coordination is particularly important because of changing task goals, time pressure and shifting team membership (Klein et al., 2006).

Multiple researchers have highlighted the necessity of teams to equally focus on both teamwork and taskwork (Lim & Klein, 2006; Marks, Mathieu, & Zaccaro, 2001; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). In order to do so, one must first understand the team task including the work of each individual team member in terms of their function and capabilities within the team (Grote, Zala-Mezö, & Grommes, 2004). In addition, the context or environment in which the team performs enhances understanding of requisite task and teamwork (Klein et al., 2006; Xiao, MAckenzie, Patey, & LOTAS Group, 1998). Different coordination mechanisms may be effective in different situations and tasks (Tschan et al., 2011a) and at different stages throughout a task (e.g. explicit vs. implicit coordination modes) (Entin & Serfaty, 1999; Serfaty, Entin, & Volpe, 1993; Xiao & LOTAS Group, 2001).

Generally, coordination requirements vary depending on characteristics of the task (e.g. task complexity, task interdependence), characteristics of the team itself such as familiarity of its members (Foushee, Lauber, Baetge, & Acomb, 1986; Kanki & Foushee, 1989), and the situation (e.g. time pressure, routine vs. non-routine procedure) (Kontogiannis & Kossiavelou, 1999).

In the literature on teamwork in complex work systems there is a growing consensus that the ability to adapt to or absorb variability, disturbance, etc. may be increased by specific team processes (Brehmer, 1996; Entin & Serfaty, 1999; Serfaty, Entin, & Deckert, 1994; Serfaty et al., 1993).

Adaptation in general includes distinct modes: a) adaptations concerning the input into the teamwork process such as a mobilization of additional resources or a structural reconfiguration of the team, and b) process adaptations, i.e. changes in coordination mechanism, decision making, and communication patterns in response to unexpected events (Burke, Stagl, Salas, Pierce, & Kendall, 2006). Although the two modes generally complement each other, the latter becomes predominant in situations where teams have limited or no access to additional resources and personnel.

Understanding the team task and the associated coordination requirements

Because of the complexity of tasks undertaken by action teams, a single point or summary assessment of coordination would be of limited utility. As these tasks are dynamic, coordination requirements inevitably change throughout a task when new information is discovered and integrated, or feedback on ongoing task strategies is given. In the case of critical events, the raised level of coordination requirements is usually traced back to need for simultaneous execution of several tasks, which is accordingly coupled with a rise in "synchronising activities, avoiding conflicts and handing over tasks" (Kontogiannis & Kossiavelou, 1999, p. 108). During these episodes, the ability to cross-monitor team member activities may be reduced. Also, more or less effort may be needed by the team at different points in the task to coordinate their work because the interdependency of the tasks that they are undertaking vary.

In teams, team members move "seamlessly in and out of synchronization with one another, allowing for both independents and varying levels of conjoint activity" (Watts & Monk, 1998) p. 1563). This means that their tasks are highly interdependent. Generally, four types of task interdependence are distinguished (Tesluk, Mathieu, Zaccaro, & Marks, 1997): 1. Pooled interdependence is when system performance is an additive function of individual performance. Coordination is usually achieved by a centrally determined work process that every team member can follow individually, but can complete in parallel to achieve a common goal. 2. Sequential interdependence is a linear workflow, where each individual's work is dependent on proper fulfilment of prior tasks. Accomplishment of each work step triggers the next work step to begin. 3. Reciprocal interdependence means that information and results of work activities have to be exchanged between the team continuously. Coordination is achieved through direct communication between team members. 4. Intensive interdependence is where team members work closely together, but workflow is poly-directional, flexible and intensive. In this case, multiple forms of coordination are necessary because the team repeatedly faces new situations that must be interpreted and solved within the team (Grote et al., 2004).

Using task deconstruction to understand interdependence and coordination

To best understand the contribution of the task to team performance and the influence the task has on coordination, task deconstruction can provide a frame of reference for analysing coordination requirements. Previous research has used task analysis to explicitly identify cognitive requirements in performing complex work in naturalistic environments (Crandall, Klein, & Hoffman, 2006; Roth, 2008). Other researchers have utilized hierarchical task analysis to generate an in depth understanding of task steps and requirements, in order to identify necessary team processes (Tschan et al., 2011b). Utilizing task analysis in complex work environments has shown that the task has significant influence over the team's coordination requirements.

In healthcare action teams, there is utility in determining generic coor-dination requirements that cross-situational boundaries. To date, researchers have examined different acute care environments in isolation, not examining potential overlaps in coordination requirements across different teams and settings. The result is multiple, idiosyncratic lists of coordination requirements. While specificity is helpful for understanding expertise, a generic set of requirements could be useful for research and training, providing a unified common language for experts to use. There may be common characteristics of effective performance that can be identified by integrating data from different action team settings. Based on research gaps identified in this literature, the goals of this paper are:

1. To identify coordination requirements within complex healthcare action team tasks using expert interviews and task analysis

2. Compare coordination requirements between clinical settings and identify potential generic coordination requirements of healthcare action team coordination

METHOD

Study setting

Three unique study task environments are included in this paper. Though the authors did not design their respective research studies specifically to compare coordination requirements in these three environments, each used similar methods to understand the clinical task and identify the associated coordination requirements and possible triggers for adaptive coordination. Further detail on each of the studies can be found in previous publications (Henrickson Parker et al., under review; Manser, Howard, & Gaba, 2006; Schmutz & Manser, 2014).

The first task environment of interest is cardiac anaesthesia. In this study, the focus was on routine clinical work for aesthetic teams, which usually consist of an anaesthesia resident (trainee) and an attending anaesthesiologist (expert). The anaesthesiologist is responsible for sedating the patient and ensuring that sedation remains adequate and the patient is stable throughout an operation. In cardiac surgery, because the heart may be stopped depending on the surgical procedure, the anaesthesiologists are also responsible for working closely with the perfusionist (a specialist who runs the heart lung machine) and the surgeons to safely go on and come off cardiopulmonary bypass.

The second task environment of interest is a simulated emergency for paediatric septic shock. This type of emergency can occur in multiple clinical environments (ICU, floor, etc). Septic shock is the condition resulting from a blood infection typically caused by bacteria. It includes extremely low blood pressure, a rapid heart rate and altered mental status and is a life-threatening situation requiring immediate attention. In this simulation scenario, teams were asked to treat a six-month-old boy. The teams had to diagnose the boy with acute septic shock and then treat him. In this situation, there is a medical emergency, but once it is diagnosed, the team has a clear treatment protocol to follow.

The final task environment of interest was trauma resuscitations. In a typical resuscitation, patients are brought into a specialized unit designed for immediate life saving treatment and rapid diagnosis and disposition. Patients are acutely ill and require immediate attention. The team has very little knowledge of the patient's needs prior to the patient arriving, and is often ad hoc, consisting of individuals that have seldom or never worked together. In this environment, the team must constantly re-evaluate the situation and update their workflow accordingly. In this study, observations were conducted in a live environment at a level 1 trauma centre.

Procedure

For each study, familiarizing observations were first conducted. This step is essential in understanding not only the specific task demands, and has been shown to yield significant benefit in safety analyses of other high risk industries (Thomas, Sexton, & Helmreich, 2004). Familiarizing observations allows the individual deconstructing the task and documenting the coordination behaviours to see beyond the extraordinary technical aspects of clinical care (e.g. open heart surgery) and focus on the team coordination behaviours.

Once familiar with the task environment, the main task (cardiac anaesthesia, paediatric septic shock care, trauma resuscitation) was deconstructed using standard task analysis techniques (Annett, 2004; Kirwan & Ainsworth, 1992) for teams (Stanton, Salmon, Guy, Baber, Jenkins, 2005). In each case, the task analysis was created based on procedural manuals, but also on iterative input from expert interviews and panels and from targeted observations complemented by conversations with the clinicians that were observed where possible. In analysing the team task we each documented the task goals and procedural steps and included instructions (e.g. do 1-4 in order, then do 5 if necessary), timeframes, interdependencies and the person or subgroups that should execute a specific procedural step.

Finally, coordination requirements for each procedural step, groups of steps, or task sequences in the task analysis were gathered from expert interviews. The focus and the specific procedure of these expert interviews are discussed in further detail for each study below.

Eliciting expert knowledge on coordination requirements

Cardiac anaesthesia. The goal of this study was to understand the coordination requirements during typical cardiac procedures as well as triggers for adaptive coordination. The analysis was centered around the anesthesia team and their coordination with each other and with other team members.

In the first step, an outline of the goals and procedural steps in cardiac anaesthesia was drafted using information extracted from guidelines on cardiac anaesthesia and field observations complemented by conversations with the clinicians that were observed. This task representation was then submitted to an expert panel consisting of four cardiac anaesthesia attendings for review.

In the second step, semi-structured interviews based on the initial task analysis were conducted to determine coordination requirements in cardiac anesthesia with all six cardiac anesthesia attendings, five third year anesthesia residents who had just completed their cardiac rotation, and the two cardiac surgeons for the participating hospital. Participants were asked to describe the coordination requirements within the anesthesia team as well as to other members of the perioperative team in terms of information exchange, task distribution etc. a) during different phases of cardiac anesthesia, b) for two different surgical procedures (i.e. bypass surgery (CPB) with and without cardiopulmonary bypass (NOCPB)), and c) how these coordination requirements change in the case of an unexpected clinical event.

Interviews were conducted by a human factors researcher familiar with the process of cardiac anesthesia (average duration 63 minutes for attending anesthesiologists, 48 minutes for residents, and 23 minutes for surgeons). Interviews were audio-recorded, transcribed verbatim and a qualitative content analysis was performed.

Pediatric Septic Shock

The goal of this study was to better understand the coordination requirements based on a team task analysis of an emergency situation with clear protocols for treatment.

To develop the task analysis, three experienced acute care paediatricians and simulation educators listed all relevant tasks based on published European Resuscitation Council guidelines and their own clinical experience. This draft was then sent out to five paediatric experts for reviewing using an adapted Delphi-Method (Gordon, 1994). After three review rounds consensus was achieved (detailed information about the procedure (Schmutz, Eppich, Hoffman, Heimberg, & Manser, *in press*). Based on video observations of teams managing septic shock in a simulated setting and further discussions with clinical educators, goals and sub goals were extracted and all tasks were organized and represented in a hierarchical manner.

In the second step, we conducted three semi-structured interviews on coordination requirements in this specific septic shock task. Each interviewee received the team task representation beforehand. Thereupon they were instructed to name the coordination requirements most important to achieving the specific sub-goals in a given phase of treatment. To provide a common basis for discussion of coordination requirements we provided descriptions of four non-technical skills that overlap with commonly used categories for coordination behavior (i.e. task management, team working, situation awareness and decision making (Flin, Patey, Glavin, & Maran, 2010).

Trauma Resuscitation

The goal of this study was to understand the task of trauma resuscitation and the associated necessary coordination behaviours for each stage within re-suscitation.

In the first step, a hierarchical task analysis (HTA) was created. The HTA was constructed by experts in human factors and trauma resuscitation and was based on standard Advanced Trauma Life Support (ATLS) guidelines.(American College of Surgeons, 2008)

In the second step, to understand coordination requirements for each goal and sub-goal, expert interviews were conducted (n=27). Handwritten notes were taken during the interviews. To utilize the expertise of the individuals involved, each participant was presented with the HTA for review prior to the interview. They were then asked if the HTA was a realistic reflection of what they did during resuscitation, and whether or not additional steps added or steps deleted or changes should be made. Interviewees were also given a list of coordination behaviours and definitions for each behaviour. They were asked to assign specific coordination behaviours to each task where appropriate. To fully understand the unique nature of how the task influences team coordination, interviewees were asked what coordination or team performance behaviours made a resuscitation “go well” and what made it go “poorly” or feel “chaotic”. The handwritten notes were transcribed, and the notations of changes were listed for each separate step in the HTA.

Comparing coordination requirements across healthcare action team settings

All three task analyses were distributed among the research team. Research team members then familiarized themselves with each analysis and a detailed discussion on similarities and differences in coordination requirements was conducted. During this discussion, the group identified instances of task interdependence that occurred during each type of healthcare action team setting. In addition, during these discussions, aspects of similarity across the different task environments were discussed. Coordination requirements and task requirements were discussed among the team, particularly focusing on the temporal and task dependence of certain coordination requirements.

RESULTS

Task structure and coordination requirements in three healthcare action team settings

For each of the three task environments, the task analysis served as a baseline for understanding the task and therefore a structure to build understanding about the variable nature of coordination requirements.

Each task analysis revealed a "gold standard" process, but with potential for marked deviation should patient or task parameters change. It was difficult to create an exhaustive list of every potential variation in the task analyses, but major deviations were easily captured. Major deviations showed an increase in coordination demands in the anaesthesia setting, but in the emergency settings (paediatric septic shock and trauma resuscitation), experts did not discuss a marked increase in coordination demands, as the demands were already fairly high.

Coordination requirements in cardiac anesthesia

In cardiac anaesthesia, experts described a "template" of the procedural steps and the parameters to be met throughout a case. Within this template, certain decisive episodes where the tasks of multiple team members are more closely linked were highlighted as requiring more and/or different kinds of coordination. For example, during the stage of the procedure where the patient is taken off cardiopulmonary bypass and the heart is restarted after a repair, the coordination requirements between the surgeon, perfusionist and anaesthesiologist increase, according to interviewees.

Any deviation from the standard work template was described as triggering increased coordination: "for most procedures that you know, you know, you have this kind of ideal course, a template in your brain [..] And then if things start deviating from that template [..] I mean I'm kind of running this program of what should be going on. I'm always comparing it to things that are happening left and right.“ (A2). Depending on the kind and degree of deviation from the expected course of action, specific coordination behaviours such as closer monitoring of the patient and other team members, providing a situation assessment to the team, discussing options or setting new priorities or reassigning tasks were discussed: "If something happens so that it's unusual, then you have to kind a step back, reevaluate all the, you know, the situation, talk with the team, the members of the team and then figure out what might be best for the patient." (A3) For recognized emergency situations the team may decide to switch to a different template altogether from which to work: "You know, if somebody's just behaved funny, I'll say, "You know, this is what I have drawn up. I've got Epi, I've got dopamine." We always have dopamine and nitro. They [surgeons] know that. We know that. The patient is acting funny. We might be doing something differently." (A3)

Participants highlighted several differences between CPB and NOCPB procedures. Specifically, a higher level of task interdependence and a resulting increase in coordination demands was discussed during "beating heart" or NOCPB cases: *"I think the off-pump stuff [NOCPB] has more issues about what the surgeons are doing and what the anaesthesiologist is doing, being coupled and really affecting each other. There's a more significant period of that."* (A6) During these cases, the anaesthesiologists described increasing internal coordination among the anaesthesia sub team but also with other team members when compared to similar stages during CPB cases. The change in task results in different coordination demands at different stages and much shorter cycles of updating if the level of coordination is still appropriate.

Coordination requirements in pediatric septic shock

In paediatric septic shock, coordination requirements were described as varying depending on the stage of the resuscitation. In this task, we identified one primary goal (survival of the patient) and four sub goals: 1) First assessment and diagnose of septic shock, 2) Primary treatment and patient stabilization, 3) Treatment and re-evaluation and 4) Reassessment and planning for other treatments (Figure 1). Once a diagnosis is made (sub goal 1), the secondary and tertiary sub goals are standard, or template. This template is frequently checked and any deviations require an updating of the template and a potential shift in goals, priorities and templates to work from.

The three experts agreed that in the primary stage of the treatment, maintaining situation awareness is the most important coordination requirement. In the later stages of the treatment, team working (e.g. supporting others, exchanging information, coordinating specific activities with team members) and task management are increasingly important. Decision-making becomes important after sub goal 1, or when the decision needs to be made as to what treatment the team should initiate. This decision needs to be communicated explicitly: *"As soon as someone comes up with the diagnosis it has to be communicated to the whole team. Differential diagnosis always has to be communicated explicitly"* (P1). Decision-making remains important until the end of sub goal 3 because there is a constant need for reassessment and updating to monitor the patient's changing hemodynamic status. After each clinical intervention, a decision needs to be made about the next steps in the process (i.e. transitions between sub-goals identified in the task analysis).

The septic shock scenario was described as a highly leader-centric scenario. Almost all information is funnelled through the leader and he or she distributes all the tasks: *"The leader is responsible for the coordination of the team. The best thing is when there is enough staff, if the leader is not included in the process and just responsible for coordination. If he has to do something he is often absorbed and cannot maintain the overview of the whole team"* (P2). In this situation it is important to clearly assign the role of the leader in advance: *"A team has to define its leader. Sometimes I think my colleague is the leader because he has more experience so I expect him to coordinate the team but at the same time he thinks I am the leader because I have more experience in this specific case. This leads to a disaster…"* (P2). The task of treating paediatric septic shock is highly interdependent. However, some actions can explicitly be performed simultaneously, or require pooled interdependence, while others require sequential or reciprocal action. Because of the various actions executed at the same time the roles in the team must be clear also for experienced teams: *"Roles have to be assigned clearly, if possible in advance. People often think it is not necessary, especially in well-established teams, but this is not true. Roles and sub teams always have to be clearly assigned before every case"*(P3). Although there are many pooled tasks that are ongoing (e.g. when nurses are preparing medication during sub-goals 2 and 3), the leader is a sequential anchor, making decisions, updating and reassessing and giving direction to the team. For example, though there are nurses preparing medication while the team is continuing its work on the patient, they do not finish their task goal (i.e. give medication) until the team leader gives them approval.

Coordination requirements in trauma resuscitation

In trauma resuscitation, experts had difficulty describing a single coordination requirement for specific task steps. Experts stated that "all" coordination behaviours could be important at each task step, depending on the patient's acuity and how many team members were involved. In this task, there is one main goal, the survival of the patient to the next step in their care, and 5 sub-goals: 1) airway maintenance with cervical spine protection, 2) breathing and ventilation, 3) circulation with haemorrhage control, 4) neurologic evaluation and 5) exposure and environmental control (ABCDE- Airway, Breathing, Circulation, Disability, Exposure). At each step, there are certain sub-goals that must be met prior to moving on to the next task step (Figure 2). Once a sub-goal is reached, rechecking is required because a patient can deteriorate at any time. Therefore, although a sub-goal is reached, it is not considered "finished".

The situation is highly influential in trauma resuscitation, so experts had difficulty discussing generic coordination requirements. There were particular points at which monitoring was necessary, and a decision would be made based on incoming treatment information. For example, in the "Circulation" assessment stage, IV fluid therapy is initiated. Then the patient's response must be observed before the next decision can be made. At this point, the team's work is highly interdependent, with much pooled work being conducted, but also work is sequentially anchored based on the patient's response to fluid. The work is also intensively interdependent at this point because the task goals could change at any point in time, and reassessments are part of the team's task.

The task of trauma resuscitation is highly leader centric, similar to paediatric septic shock. In the two hospitals included in our sample, the team leader, or the individual touching the patient and conducting the assessment is usually a learner. Therefore, though they serve as a functional funnel through which most procedure related tasks must run, there is also a more expert consultant level physician (attending) who is usually maintaining awareness of the whole team simultaneously. Ultimately, the attending will make most of the treatment decisions but they will be executed through the team, requiring significant explicit coordination. The attending serves as a coordination funnel, allocating most tasks and monitoring team progress.

Though standard templates exist (e.g. ABCDE mnemonic) for resuscitation, within each template there are a number of criteria that are constantly being assessed by the team. Because the team has limited information on the patient and their condition prior to seeing the patient, experts stated that their coordination is especially important during the first few minutes of the patient's arrival in the trauma unit. In this period, there is intense task interdependence, involving most of the team members working in sub teams, accomplishing tasks simultaneously and sequentially. During this time, multiple templates are considered and rejected or updated according to the patient's parameters. Throughout the trauma resuscitation, coordination requirements are highly variable because of the presentation of new or updated information.

Characteristics of coordination requirements across healthcare action team settings

Three overall themes emerged from the group discussions pertaining to coordination requirements in healthcare action teams. General coordination requirements include 1) continual reassessment, 2) making coordination needs explicit based on monitoring "anchors", and 3) the occurrence of a non-routine situation requires explicit coordination.

Continual Reassessment. In healthcare action teams, the task and the associated goals and sub goals may range from rather stable (cardiac anaesthesia) or highly dynamic (trauma). However, the stability of the previous minute does not indicate the requisite level of coordination for the next. Constant reassessment of coordination requirements is necessary for smooth and effective task performance. For example, though the task of cardiac anaesthesia is rather template-driven, at any given point, the patient could become unstable or the surgical environment could change, and the situation may require reassessment and an updating of coordination requirements. In paediatric septic shock, the situation is initially dynamic, but once a diagnosis is made, the task requirements are well defined, so coordination requirements are more evident once the initial diagnosis is made. However, at any point, the patient could become unstable and a different intervention may be necessary. In trauma resuscitation, the team is learning more about their task as they are engaged in it. For example, the patient may enter the trauma bay labelled as a "pedestrian hit by a car" but upon visualization of an emergency CT scan, it could be seen that the patient was actually shot, thus completely changing the next steps for the team. Therefore, the initial template of "blunt trauma, possible internal injury" must be updated to "penetrating injury to the brain" which radically changes the clinical task. This change has to be communicated across the team and will also alter the associated coordination requirements.

These examples highlight how teams in all three task environments have to adapt their coordination between episodes requiring effective management of task related information (e.g. initial assessment and diagnosis) or of task execution (e.g. giving fluids).

Making coordination needs explicit based on monitoring "anchors". In cardiac anaesthesia this was particularly evident at certain points during the procedure, which we labelled as "anchors". At these anchor points, there were particular task characteristics that were indicative of the need for a different type of coordination. For example, the attending anaesthesiologists would indicate to the trainee that they were leaving the room briefly, but that they should be called as soon as the surgeons prepared for a certain procedural step. This request served as an anchor for the trainee in the room, and showed that with changes in task interdependence coordination requirements would change. In addition to these anchors inherent in the team task, team members may define additional anchors based on changes in clinical conditions or their anticipation of increased need for re-evaluation and decision making for a specific patient.

In paediatric septic shock and trauma resuscitation, these anchors were also present. Similar to anaesthesia, these anchor moments were significant clinical moments (such as transitions between sub goals; Figures 1 and 2) or points at which the patient parameters would indicate a different level of intervention was necessary (e.g. the child losing consciousness). In these emergency scenarios, anchors were closely associated with significant clinical decisions.

Non-routine events require explicit coordination. In all three task environments non-routine events were described as requiring a shift to more explicit coordination; if only to indicate a shift from one template to another. In cardiac anaesthesia all team members possess an in-depth understanding of interrelations between the tasks of the various sub-teams and coordinate implicitly over extended periods of a case. However, all of them were expected to make any deviations explicit so that the team can re-evaluate and decide on how to proceed and potentially define additional anchors for explicit coordination. The septic shock scenario starts with routine tasks (sub goal 1; e.g. equipment check, connect monitors, ABC assessment) and coordination is rather implicit because the team is aware of the template for the initial assessment. The moment when the patient loses consciousness and the vital signs drop indicates a non-routine event that impacts coordination requirements. Because of the patients' critical state, the team has to act quickly and communication is mostly explicit (e.g. stating decisions, giving orders, distributing tasks). As soon as the patient is stable (sub goal 4) there is a shift again to more implicit coordination within a template for treating the septic shock (e.g. nurses preparing the next fluid bolus without explicit instruction). Trauma resuscitation provides a unique environment to examine non-routine events and their impact on team coordination. Because the task of trauma resuscitation involves multiple templates that are constantly updated and reassessed, identifying instances of non-routine events is difficult via task analysis. It may be that non-routine events are actually considered routine by clinicians working in this task environment, because they are so common. Overall, coordination in this task environment is much more explicit while coordination within sub-teams may be implicit within a specific procedural step or as long as the template does not change. However, there is a recognized need for constant explicit updating to assure effective overall coordination. Because this is rather specific to trauma resuscitation a specific role has been defined within these teams to funnel the excessive amount of explicit coordination.

In terms of adaptive coordination, these examples refer to the necessary shift between explicit and implicit forms of coordination. Across all three study settings experts stated that when confronted with a non-routine event, there were opportunities for coordination to be made explicit and that missing these opportunities is detrimental to team performance. It is important to note that while the need for adaptive coordination on the explicit-implicit dimension was described as driven by situational changes and non-routine events, the need for adaptive coordination on the information-task management dimension was more clearly linked to the task structure.

DISCUSSION

The goal of this paper was two-fold. First, we identified coordination requirements within the complex task environments of healthcare action teams by eliciting expert knowledge using a task analysis approach. Second, we compared coordination requirements across clinical settings and identify common characteristics of coordination requirements in healthcare action teams that can inform future research and interventions aimed at improving team performance.

We identified three different working modes for healthcare action teams: template-based work, work that must be conducted according to an evolving template, and finally, work that is based on templates that only include a set series of steps until the next clinical indicator is evident. In each situation, the team starts with an initial "template" of actions, which is similar to the concept of initial coordinateness (Wehner, Clases, & Bachmann, 2000). The template must be updated regularly to integrate new information and potentially adjust for non-routine events, which require the team to engage in coordination activities and move to a new form of coordinateness (Wehner et al, 2000). In cardiac anaesthesia, the template is usually fairly accurate, with significant but limited deviations that would cause a shift in coordination requirements and a potential shift to another template. In paediatric septic shock, once a diagnosis has been made, the template is fairly clear, but again may need to be updated with incoming information (e.g. patient does not respond to initial fluid bolus). In this environment, the team is initially seeking a template for action, and once they have established a diagnosis, they can use the associated treatment protocol as a template and update as appropriate. In trauma resuscitation, the team is constantly developing and choosing templates as they discover new information. A template is chosen for one step, and that template may be confirmed or refuted within moments (e.g. blood pressure is dropping, so fluid is given, but does not result in increase pressure, so a new assessment must be made).

There are a repertoire of patterns that expert practitioners use when confronted with complex work situations (Klein, 2008). Research examining expert performance in other high risk settings has shown that experts make extremely rapid situation assessments, followed by task-related decisions that have a significant impact on the course of action and the associated coordination requirements for the team (Klein, 2008). The exact nature of the resulting changes in coordination requires further research. In each of our study environments, experienced team members were able to rapidly recognize situations that are typical and those that are atypical. Our results highlight that expert team performance requires both depth of understanding to choose templates for performance, and anticipation of changes and updating to make sure the team is able to adapt to evolving situation requirements (Klein, 1989).

Methods to elicit expert knowledge have focused on utilizing anchors to help experts describe processes (Ford & Sterman, 1998) both graphically and through describing situations. In our research, we found that experts used "anchors" as points at which there was potential for changing coordination requirements. Anchors served as more than process descriptors. They were triggers for the team leader or a sub team to understand that the nature of coordination must adapt at a particular point in the task. This finding is similar to the conceptualization of anchors in distributed task teams (McNeese, Theodorou, Ferzandi, Jefferson Jr., & Ge Jr., 2002). Although our healthcare action teams were all collocated, it may be that because the teams are made up of multiple sub teams, their shared cognition functions are similar to those of distributed teams.

Task interdependence is a key characteristic of team tasks to be considered when studying coordination. However, our study revealed that this relationship is even more complex in healthcare action teams. Initially, we sought to define task stages across all three tasks that are characterised by certain types of task interdependence, thus posing specific coordination demands for the team. We found that because of the dynamic requirements of the clinical task and the design of the teams, identifying consistent moments of task interdependence was difficult. Teams are designed to be flexible, but in large action teams, this may mean that they are organized as teams of teams with potentially fluid membership and roles, and sub teams that form and dissolve depending on task requirements (Klein et al., 2006). At any point, some tasks within the team may be sequential, while some may be pooled, some intensive or some reciprocal. Teams can reconfigure at multiple points, involving new members, or replacing others (e.g. in trauma an x-ray technician may be called and involved in clinical care for a short period). Team membership and task allocation are often conducted on an “as needed” basis, resulting in a constant shifting of multiple overlapping task interdependencies within the overall team.

Although standardization has been documented as important for limiting uncertainty in teams (Grote, Zala-Mezö, & Grommes, 2003), action teams require flexibility and adaptive expert performance. Nevertheless, at periods during an action team task, standardized coordination behaviours may be appropriate (Zala-Mezo, Wacker, Kunzle, Bruesch, & Grote, 2009). Given these findings we conclude that it is ultimately not possible to determine the coordination necessary for exceptional performance based on a task analysis alone because the situational variations have to be taken into account as well.

Possible ways to optimize coordination

Our work has shown that coordination is highly task specific. However, it is also necessary to consider situational variations to determine if coordination was “appropriate”. The opportunity to train teams in specific coordination behaviours based on the requirements of their clinical task is critical for healthcare teams. As shown in other high risk industries, training on team coordination and adaptation can help to improve coordination performance (Espevik, Helge Johnsen, & Eid, 2011). Training may use the heuristic of task interdependence (and multiple different task interdependencies that can coexist in the team at the same time) to train for adaptive coordination.

The leader is a very important part of coordination in healthcare action teams. The leader serves as a task “funnel”, taking in information and task requests from the team, synthesizing them and updating them based on current situation requirements, and then distributing the tasks to individuals or sub teams. Therefore, if tasks are done at the same time the team has to share the collected knowledge, the leader can serve as a conduit for the team. Leadership is closely linked with team performance and especially in task distribution, a critical coordination requirement (Cooper & Wakelam, 1999; Xiao, Seagull, Mackenzie, Ziegert, & Klein, 2003). Training for team leaders focused on optimal coordination structures may be appropriate.

In addition, teaching leaders to recognize changes in the situation and how to adapt coordination appropriately may be helpful. Changes in a patient's course which can be identified through procedural steps, care episodes or by time markers can serve as anchors for indicators of changing coordination requirements. Identifying when these anchors exist, gives clinicians the opportunity to explicitly understand that requirements for coordination are changing, and thus they must change their behaviours. Other researchers have found that explicit identification of decision points is necessary, especially for learners (Moulton, Regehr, Mylopoulos, & MacRae, 2007). If the team is not aware that they are passing a significant point in the operation, they miss the opportunity for adaptation all together and this has a negative effect on performance.

Implications for moving naturalistic decision making research forward

This study highlights the benefits of comparing coordination requirements across settings. To our knowledge, this is the first study that analysed and compared coordination requirements across routine and emergency situations.

The combination of these three studies has shown that it may be inappropriate to determine coordination behaviour ex ante, or that it can only be discussed to a certain degree by using a task analysis. Coordination requirements are highly situational dependent. This finding will require refined analytic approaches that are able to include documentation of temporal variability and task variability. Researchers must also recognize the importance of coordination anchors and templates for team performance. Anchors and templates must be further researched, because there are significant implications for training of both novice and expert clinicians utilizing these mechanisms.

We found that the level of task interdependence is highly influential on how the team must coordinate their actions. The issue of multiple overlapping and dynamically changing levels of task interdependence is an issue that, to our knowledge, has not been addressed explicitly by previous research and provides a highly challenging but potentially fruitful field of study.

Limitations

There are a few limitations to this study. First, as with all studies eliciting expert knowledge of a complex task our study is based on the experience of experts and does not provide any empirical data linked with team performance. The quality of the data was thus dependent on experts' ability to verbalise their understanding of the issue under discussion. Each of our three studies found that it was difficult for experts to use "coordination language" but not difficult to understand the concept of coordination, or to describe its importance to team performance. The reason for this could be three-fold: first, as most experts, clinicians often frame and discuss their work in terms of decision trees, rather than as a set of hierarchically arranged goals. However, within a task analysis, it is challenging for such a highly dynamic task to exhaustively discuss every possible situation and the associated coordination requirements. Second, it may be difficult for experts to discuss coordination because coordination requirements are not stable across an entire interaction. There is a continual influence of context and a holistic pattern for the entire care episode is inappropriate. A third possible reason is that coordination is so deeply ingrained in the teams, especially implicit forms of coordination, that they are unable to articulate their performance in a general sense (Klein, 1997). Task analysis must remain rigorous but avoid rigidity of method in order to maximize expert knowledge (Hoffman, Crandall, & Shadbolt, 1998).

Second, we did not initially set out to combine the three task analyses, and therefore our methods, though consistent were not identical. However, the use of an established methodological approach - familiarization with environment, constructing a task analysis, and utilizing expert interviews to understand coordination behaviours is consistent. Each of our task analyses is a representation of the optimal treatment procedure. Teams act differently in real life, and a prescribed treatment path is rarely present. Task analysis is limited in its ability to capture the complexity of environmentally driven changes, while remaining usable. Further to this point, the level of detail of the tasks needs to be defined prior to conducting a task analysis. A useful resolution for the subtasks needs to be applied. The three studies presented used a similar level of detail to have a manageable task representation. An exhaustive list of clinical permutations was too varied to be included in each representation. Third, the three study settings differed in many ways that we discuss in our paper concerning their impact on coordination requirements. Future studies will have to validate these findings by including more healthcare action team settings; with clusters of team tasks that can are similar according to the characteristics of the task environment.

CONCLUSION

Our cross-analysis of these three task environments revealed that coordination in expert healthcare action teams is varied and requires constant updating. Experts utilize "anchors" to aid in their understanding of the situation and the coordination requirements. It is possible that these anchors could be captured and taught to trainees. If teams are able to realize that they are entering a phase of work that may be more risky than their current state, they could explicitly adapt their coordination, maintaining optimal performance, though situational demands have changed. Future research should continue to consider the necessity of establishing a detailed understanding of task structure and its influence on coordination requirements to help teams self-regulate and adapt. Utilizing real time feedback on coordination requirements could help teams to optimize performance immediately. At any point in time during clinical care, an infinite number of possibilities exists that may influence the requirements for coordination. It is in these improvisations that true expert performance resides.

REFERENCES

American College of surgeons, Committee on Trauma. (2008). *Advanced Trauma Life Support: ATLS* (8th ed.). American College of Surgeons.

Annett, J. (2004). Hierarchical Task Analysis. In H. Stanton, N.A., Hedge, A., Brookhuis, K., Salas, E., Hendrick (Ed.), *Handbook of Human Factors and Ergonomics Methods*. Boca Raton, FL: CRC Press.

Arora, S., Aggarwal, R., Sirimanna, P., Moran, A., Grantcharov, T., Kneebone, R. L., & Darzi, A. W. (2011). Mental practice enhances surgical technical skills: a randomized controlled study. *Annals of Surgery*, *253*(2), 265–270. doi:10.1097/SLA.0b013e318207a789

Bigley, G. A., & Roberts, K. H. (2001). The Incident Command System: High Reliability Organizing for Complex and Volatile Task Environments. *Academy of Management Journal*, *44*, 1281–1300.

Brannick, M., & Prince, C. (1997). An overview of team performance measurement. In M. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and Measurement*. Mahwah, NJ: Lawrence Erlbaum Assoc Inc.

Brehmer, B. (1996). Dynamic and distributed decision making. *Journal of the First Service College*, *1*(2), 17–36.

Burke, C. S., Salas, E., Wilson-Donnelly, K., & Priest, H. (2004). How to Turn a Team of Experts into an Expert Medical Team: Guidance from the Aviation and Military Communities. *Quality & Safety in Health Care*, *13(Suppl 1*, i96–i104.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology*, *91*(6), 1189–1207. doi:10.1037/0021-9010.91.6.1189

Burtscher, M. J., Wacker, J., Grote, G., & Manser, T. (2010). Managing Nonroutine Events in Anesthesia : The Role of Adaptive Coordination. *Human Factors*, *52*(2), 282–294. doi:10.1177/0018720809359178.

Cooper, S., & Wakelam, A. (1999). Leadership of resuscitation teams: "Lighthouse Leadership." *Resuscitation*, *42*(1), 27–45.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. The MIT Press.

Edmonson, A. (2003). Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies*, *40*(6), 1419–1452.

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors*, *41*(2), 312–325.

Espevik, R., Helge Johnsen, B., & Eid, J. (2011). Outcomes of shared mental models of team members in cross training and high-intensity simulations. *Journal of Cognitive Engineering and Decision Making*, *5*(4), 352–377.

Flin, R., Patey, R., Glavin, R., & Maran, N. (2010). Anaesthetists non-technical skills. *British Journal of Anaesthesia*, *105*(1), 38–44.

Ford, D. N., & Sterman, J. D. (1998). Expert knowledge elicitation to improve formal mental models. *System Dynamics Review*, *14*(4), 309–340.

Foushee, H., Lauber, J., Baetge, M., & Acomb, D. (1986). *Crew factors in flight operations: III. The operational significance of exposure to short-haul air transport operations. NASA Technical Memorandum 88322. NASA Technical Memorandum.* Washington DC.

Gordon, T. J. (1994). The Delphi Method: Futures Research Methodology. AC/UNU Milennium Project.

Grote, G., Zala-Mezö, E., & Grommes, P. (2003). Effects of standardization on coordination and communication in high workload situations. In R. Dietrich (Ed.), *Communication in High Risk Environments* (pp. 127–154). Hamburg, Germany: Buske.

Grote, G., Zala-Mezö, E., & Grommes, P. (2004). The effects of different forms of coordination on coping with workload. In R. Dietrich & T. M. Childress (Eds.), *Group Interaciton in High Risk Environments* (1st ed., pp. 39–54). Burlington, VT: Ashgate Publishing Co.

Helmreich, R. L. (2000). On error management: lessons from aviation. *BMJ*, *320*(7237), 781–785.

Henrickson Parker, S., Midgley, K., Weaver, M. L., Fairbanks, R. T., Sava, J., & Burd, R. S. (n.d.). A reliable real-time observation tool to measure team coordination behavior during trauma resuscitation. *In Prep*.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis. *Human Factors*, *40*(2).

Hunziker, S., Tschan, F., Semmer, N. K., & Marsch, S. (2013). Importance of leadership in cardiac arrest situations: from simulation to real life and back. *Swiss Medical Weekly*, *143*(April), w13774. doi:10.4414/smw.2013.13774

Kanki, B. G., & Foushee, H. (1989). Communication as group process mediator of aircrew performance. *Aviation, Space, and Environmental Medicine*, *60*(5), 402–410.

Kirwan, B., & Ainsworth, L. K. (1992). *A Guide to Task Analysis*. (B. Kirwan & L. K. Ainsworth, Eds.). London: Taylor and Francis.

Klein, G. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), (Vol. 5, pp. 47–92). New York: JAI Press.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. *Naturalistic Decision Making*, 285–292.

Klein, G. (2008). Naturalistic decision making. *Hum Factors*, *50*(3), 456–460.

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. (2006). Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly*, *51*, 590–621.

Klinger, D., & Thorsden, M. (1998). Team CTA applicaitons and methodologies. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 206–209).

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To Err Is Human - Building A safer Health System*. (J. M. C. Linda T.

Kohn and Molla S. Donaldson, Ed.) (p. 287). Washington, D.C.: National Academy Press.

Kolbe, M., Burtscher, M., Manser, T., & Barbara, K. (2011). The Role of Coordination in Preventing Harm in Healthcare Groups : Research Examples from Anaesthesia and an Integrated Model of Coordination for Action Teams in Health Care, 75–92. doi:10.1007/978-3-642-15355-6

Kontogiannis, T., & Kossiavelou, Z. (1999). Stress and team performance: principles and challenges for intelligent decision aids. *Safety Science*, *33*(3), 103–128. doi:10.1016/S0925-7535(99)00027-2

Lim, B., & Klein, K. J. (2006). Team mental models and team performance : A field study of the effects of team mental model similarity and accuracy, *418*(October 2004), 403–418.

Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R., … Grober, E. (2004). Communication failures in the operating room: An observational classification of recurrent types and effects. *Quality & Safety in Health Care*, *13*(5), 330–334. doi:10.1136/qhc.13.5.330

Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiologica Scandinavica*, *53*, 143–151. doi:10.1111/j.1399-6576.2008.01717.x

Manser, T., Harrison, T. K., Gaba, D. M., & Howard, S. K. (2009). Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesthesia and Analgesia*, *108*(5), 1606–15. doi:10.1213/ane.0b013e3181981d36

Manser, T., Howard, S., & Gaba, D. (2008). Adaptive coordination in cardiac anaesthesia: a study of situational changes in coordination patterns using a new observation system. *Ergonomics*, *51*, 1153–1178.

Manser, T., Howard, S. K., & Gaba, D. (2006). Self -regulation as a central mechanism to collaboratively manage unexpected events in complex work environments. In *13th European Conference on Cognitive Ergonomics*. Zurich, Switzerland.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356–376.

Marks, M. A., & Panzer, F. J. (2004). The influence of team monitoring on team processes and performance. *Human Performance*, *17*(1), 25–41.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273–283.

McNeese, M. D., Theodorou, E., Ferzandi, L., Jefferson Jr., T., & Ge Jr., X. (2002). Distributed cognition in shared information spaces. In *Proceedings of the Human Factors and Ergonomics Society* (pp. 556–560).

Mishra, A., Catchpole, K., Dale, T., & McCulloch, P. (2008). The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surgical Endoscopy*, *22*(1), 68–73. doi:10.1007/s00464-007-9346-1

Moulton, C. E., Regehr, G., Mylopoulos, M., & MacRae, H. M. (2007). Slowing Down When You Should: A New Model of Expert Judgment. *Academic Medicine*, *82*(Suppl). doi:10.1097/ACM.0b013e3181405a76

Naikar, N., Pearce, B., Drumm, D., & Sanderson, P. M. (2003). Designing teams for first-of-a-kind, complex systems using the initial phases of cognitive work analysis: case study. *Human Factors*, *45*(2), 202–217.

Riethmuller, M., Fernandez Castelao, E., Eberhardt, I., Timmermann, A., & Boos, M. (2012). Adaptive coordination development in student anaesthesia teams: a longitudinal study. *Ergonomics*, *55*(1), 55–68. doi:10.1080/00140139.2011.636455

Roth, E. M. (2008). Uncovering the requirements of cognitive work. *Hum Factors*, *50*(3), 475–480.

Schmutz, J., Eppich, W. J., Hoffman, F., Heimberg, E., & Manser, T. (n.d.). Five steps to develop checklists for evaluating clinical performance: An integrative approach. *Academic Medicine*.

Schmutz, J., & Manser, T. (2013). Do team processes really have an effect on clinical performance? A systematic literature review. *British Journal of Anaesthesia*, *110*(4), 529–544.

Schmutz, J., & Manser, T. (2014). Using hierarchical task analysis to identify different coordination requirements in a pediatric emergency scenario. In *Poster presented at the International Symposium on Human Factors and Ergonomics in Healthcare, Chicago, USA*. Chicago, IL.

Serfaty, D., Entin, E. E., & Deckert, J. (1994). Implicit coordination in command teams. In A. H. Lewis & I. S. Lewis (Eds.), *Science of Command and Contro: Part III: Coping with Change* (pp. 87–94). Fairfax, VA: AFCEA Press.

Serfaty, D., Entin, E. E., & Volpe, C. (1993). Adaptation to stress in team decision-making and coordination. In *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (pp. 1228– 1232).

Stanton, N A, Salmon, P M, Guy, H W, Baber, C, Jenkins, D. P. (2005). *Human Factors Methods: A Practical Guide for Engineering and Design* (p. 571). Burlington, VT: Ashgate Publishing, Ltd.

Sundstrom, E., De Meuse, K. P., & Futrell, K. (1990). Work teams: Applications and effectiveness. *American Psychologist*, *45*(2), 120–133.

Tesluk, P., Mathieu, J. E., Zaccaro, S. J., & Marks, M. A. (1997). Task and Aggregation Issues in the Analysis and Assessment of Team Performance. In M. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and Measurement* (pp. 197–224). Mahwah, N.J: Lawrence Erlbaum Assoc Inc.

Thomas, E. J., Sexton, J. B., & Helmreich, R. L. (2004). Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. *Quality & Safety in Health Care*, *13 Suppl 1*, i57–64.

Tschan, F., Semmer, N. K., Vetterli, M., Gurtner, A., Hunziker, S., & Marsch, S. U. (2011a). Coordination in Human and Primate Groups, 93–115. doi:10.1007/978-3-642-15355-6

Watts, L. A., & Monk, A. F. (1998). Reasoning about tasks, activities and technology to support collaboration. *Ergonomics*, *41*(11), 1583–1606.

Wehner, T., Clases, C., & Bachmann, R. (2000a). Co-operation at work: a process-oriented perspective on joint activity in inter-organizational relations. *Ergonomics*, *43*(7), 983–997.

Wehner, T., Clases, C., & Bachmann, R. (2000b). Co-operation at work: a process-oriented perspective on joint activity in inter-organizational relations. *Ergonomics*, *43*(7), 983–997.

Weick, K. E. (1993). The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly*, *38*(4).

Wiegmann, D. A., Eggman, A. A., Elbardissi, A. W., Parker, S. H., & Sundt 3rd, T. M. (2010). Improving cardiac surgical care: a work systems approach. *Applied Ergonomics*, *41*(5), 701–712. doi:S0003-6870(09)00168-9 [pii] 10.1016/j.apergo.2009.12.008

Xiao, Y., & Group, L. (2001). Understanding coordination in a dynamic medical environment: Methods and results. In M. McNeese, E. Salas, & M. Endsley (Eds.), *New Trends in Collaborative Activities: Understanding System Dynamics in Complex Environments* (pp. 242–258). Santa Monica, CA: Human Factors and Ergonomics Society.

Xiao, Y., MAckenzie, C. F., Patey, R., & Group, L. (1998). Team coordination and breakdowns in real-life stressful environment. *Proceedings of the Human Factors and Ergonomics Society*, *42*, 186–190.

Xiao, Y., Seagull, F. J., Mackenzie, C., Ziegert, J., & Klein, K. J. (2003). Team communication patterns as measure of team processes: Exploring the effects of task urgency and shared team experience. *Proceedings of the Human Factors and Ergonomics Society*, *47*, 1502–1506. doi:10.1177/154193120304701228

Xiao, Y., Seagull, J., Mackenzie, C. F., Klein, K., & Ziegert, J. (2002). Adaptation of team structure of trauma resuscitation teams . *Proceedings of the Human Factors and Ergonomics Society*, *46*, 569–573.

Zala-Mezo, E., Wacker, J., Kunzle, B., Bruesch, M., & Grote, G. (2009). The influence of standardisation and task load on team coordination patterns during anaesthesia inductions. *Qual Saf Health Care*, *18*, 127–130. doi:10.1136/qshc.2007.025973

5.4 Study D: Effective team coordination in emergency care: The moderating role of task type

Jan Schmutz[1], Florian Hoffmann[2], Ellen Heimberg[3], Tanja Manser[1]

[1] Industrial Psychology and Human Factors, Department of Psychology, University of Fribourg, Fribourg, Switzerland.
[2] Dr. von Hauner University Children‘s Hospital, Munich, Germany.
[3] University Children‘s Hospital, Tübingen, Germany

This manuscript was submitted for publication and is under revision at the European Journal of Work and Organizational Psychology

INTRODUCTION

Effective coordination is a main characteristic of high performing teams (Rico, Sanchez-Manzanares, Gil, & Gibson, 2008). According to McGraths (1984) Input-Process-Output (IPO) model coordination can be seen as a team process that converts inputs (e.g. team or task characteristics) into an output (Reader, Flin, Mearns, & Cuthbertson, 2009). During the team process, information as well as task execution constantly needs to be coordinated to ensure smooth team functioning (Kolbe, Burtscher, Manser, Kunzle, & Grote, 2010; Manser, Howard, & Gaba, 2008) and to prevent breakdowns in the quality of teamwork (e.g. communication errors, unclear division of work roles). Team coordination has been defined as the process involving the use of behaviour patterns and strategies aimed at integrating and aligning actions, knowledge, and objectives of interdependent members in order to achieve a common goal (Arrow, McGrath, & Berdahl, 2000; Brannick & Prince, 1997). In an inclusive model of group coordination Boos, Kolbe and Strack (2010) extended the rather simple IPO model and describe individual goals, meanings and behaviours as inputs that are coordinated through implicit or explicit coordination mechanisms. Further they add a temporal structure and describe pre- in- and post-process coordination that result in outputs. In the present study we focus solely on in-process coordination and its mechanisms during the team process.

In the last two decades coordination has increasingly been investigated in the context of safety management in high-risk organizations in aviation (Grote, Kolbe, Zala-Mezo, Bienefeld-Seall, & Kunzle, 2010), healthcare (Manser, Harrison, Gaba, & Howard, 2009) or the nuclear industry (Waller, Gupta, & Giambatista, 2004). Research on coordination in these high-risk environments is essential because a) errors caused by poor coordination may lead to serious consequences involving harming or even killing people and b) high risk organizations are dealing with complex team structures. People are working in interprofessional action teams (IAT) in which team members with different background, skills and roles have to coordinate their actions in intense and unpredictable situations (Edmondson, 2003; Sundstrom, De Meuse, & Futrell, 1990). Typical IAT are emergency medical teams. These teams usually consist of one experienced senior physician (leader), one or two less experienced residents and several nurses. All team members have different backgrounds, education and experience. In emergency situations the patient is in a life threatening condition and the team has to act quickly and provide the correct treatment. Thus, in these situations coordination and leadership are essential to safe and effective performance (Hunziker et al., 2010).

Various studies have linked coordination with performance (Fernandez Castelao et al., 2011; Schmutz & Manser, 2013). But coordination is not a static process. One coordination behaviour can be effective in one situation but not in an other (Grote et al., 2010). IAT in complex work environments need to adjust their coordination mechanisms to the task requirements (Tschan, Semmer, Nägele, & Gurtner, 2000), coordination has to be adaptive.

Adaptive coordination

IAT are mostly working in complex, unstable and unpredictable environments so the team has to dynamically adapt to characteristics of the task (e.g. level of standardization, complexity), characteristics of the team (e.g. familiarity of its members) or to characteristics of the situation (e.g. routing vs. non-routine phase; Manser et al., 2008). Adaptive coordination is defined as the process enabling a team to use information gathered from the task environment to adjust strategies through the use of compensatory behaviours and reallocation of intrateam resources (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995).

Burke et al. (Burke, Stagl, Salas, Pierce, & Kendall, 2006) present a comprehensive model of adaptive team performance. They describe an adaptive cycle consisting of situation assessment, plan formulation, plan execution and team learning during which teams adapt these process behaviours in response to a salient cue or cue stream. The central coordination processes in this model are defined as mutual monitoring, communication, back-up behaviour and leadership. The team has to detect cues or set of cues (e.g. task characteristics) signalling a change on the team level (e.g. coordination behaviour) that impacts performance. The whole process is influenced by inputs that are individual characteristics (e.g. team expertise), job design characteristics but also by emergent states (e.g. team situation awareness, psychological safety, shared mental models) that serve as proximal outcomes as well as inputs to the cycle. This model makes clear that team adaptation is a complex process and is influenced by many variables. The fundamental idea though is that teams have to recognize and interpret specific cues and then adapt their coordination behaviour according to the situational requirements according to these cues.

Several studies link adaptive coordination in IAT with performance. Entin and Serfaty (1999) showed performance improvements in navy teams through adap-tation training focusing on the shift from explicit to implicit coordination in the right time and to choose strategies that are effective during periods of high stress. Grote et al. (2010) showed that aviation teams with better strategies for adapting coordination to different levels of standardisation and task load achieved better performance. Burtscher et al. (2011) showed that high performing anaesthesia teams increase the amount of information management in response to a critical event. Riethmüller et al. (2012) investigated teams of medical students training and debriefing several scenarios according to CRM principles (Rall, Gaba, Howard, & Dieckmann, 2005) over a longer period of time and found an increase of adaptive coordination (i.e. shift from explicit to implicit) over time. These results imply that effective teams adapt their coordination behaviour to specific situational factors.

It can be assumed that teams not only have to adapt to specific situations within a task but different tasks also place different coordination requirements on the team (Parker, Schmutz, & Manser, under review). There may be more general behaviours like planning or overall communication that are linked with good performance in different kinds of tasks. But Tschan et al. (2000) state that these behaviours just increase the possibility of the team successfully identifying task requirements and thus lead to the task relevant behaviours that determine performance. This means there are generally effective coordination behaviours (e.g. planning) as well as the more powerful task related coordination behaviours (Tschan et al., 2000).

The question is what kind of coordination behaviour is effective for what kind of task. Studies have identified various task specific coordination behaviours. Grote et al. (2010) found that implicit coordination (e.g. anticipate actions) is effective in standardized tasks and leadership and heedful interrelating (i.e. constantly reconsidering the effects of team members own actions) are effective in less standardized tasks. Other researchers (Zala-Mezo, Wacker, Kunzle, Bruesch, & Grote, 2009) observed more implicit coordination, less leadership and less heedful interrelating in standardized phases. Further the authors observed more heedful interrelating in high task-load situations compared to low task-load situations. Tschan, Semmer, Vetterli, et al. (2011) found that in diagnostic tasks, where the main goal is to establish a diagnosis, explicit information sharing and collecting and talking to the room (i.e. commenting on one's owns actions, thinking out loud) was linked to performance whereas in resuscitation tasks, where the goal is an algorithm-based execution of the resuscitation, clear and direct leadership was most effective. Fernandez-Castelao et al. (2011) also investigated resuscitation tasks and were able to link pre- and in-process planning as well as task distribution by the leader with clinical performance. Manser et al. (2009) showed that high performing anaesthesia teams in a crisis situation exhibited less task distribution and more situation assessment than low performing teams. Burtscher, Wacker, Grote, & Manser (2010) also investigated anaesthesia teams during phases of an increased amount of non-routine events and found that more task management was positively related with performance. In another study Burtscher et al. (2011) found information management to be effective in crisis situations.

The recent literature on effective coordination in IAT investigates numerous different coordination behaviours in the context of different tasks. First, it appears there is no consensus what specific coordination behaviour is effective in what situation. There are even contradicting results indicating, for example, that more task management is effective in one crisis situation while in another less task management is effective (Burtscher et al., 2010; Manser et al., 2009). Second, no common taxonomy of tasks is used across studies making it hard to compare the results of the respective studies.

Typical task classifications are for example routine vs. complication phase (Burtscher et al., 2010; Riethmüller et al., 2012), standardized vs. non-standardized tasks (Grote et al., 2010) or low-moderate-high workload tasks (Waller et al., 2004). These task classifications seem straightforward, are obvious for the investigated tasks but are often not applicable to other tasks. For many medical emergencies no clear phases exist (e.g. routine or non-routine phases). As soon as an emergency situation in a ward occurs the healthcare worker detecting the incident calls for help and assembles the emergency team then dealing with the critical condition of the patient. In this study we focus on the emergency situation that can be seen as a non-routine phase as a whole, because usually these types of events are rare for most teams. An other classification of tasks is needed to investigate task related coordination in emergency teams.

The present study

While various coordination behaviours have been linked with performance in different tasks, no consensus exists about what specific coordination behaviour is effective in what task. We assume and the presented literature indicates that not all coordination behaviours are effective in all kind of tasks. Thus investigating coordination without taking into account the task would lead to inappropriate results. A strict use of a task taxonomy and investigating its related coordination requirements leads to more structured research and helps to formulate more specific hypothesis about coordination related to different tasks. Therefore the goal of this study is to identify task specific coordination behaviours in relation with a task taxonomy on the example of emergency medicine teams.

Taxonomy of tasks

To identify and interpret task-related effective coordination, researchers need a profound understanding of the task. This has been done by two studies in healthcare using team task analysis (Parker et al., under review; Tschan, Semmer, Vetterli, et al., 2011). The taxonomy must be able to group different emergency tasks and to reflect the specific characteristics of each task.

We propose a psychologically driven approach to categorize tasks based on Rasmussen's model of human performance (Rasmussen, 1983) that distinguishes three levels of performance of skilled human operators. *Skill-based behaviour* includes sensory-motor performance that is automated and takes place without conscious effort. *Rule-based behaviour* is typically controlled by a stored rule or procedure that may have been derived during previous experiences. Performance is goal-oriented but structured by "feed-forward" control through stored patterns of actions. The highest level is *knowledge-based* performance where based on analysis of the environment goals are explicitly formulated and the internal structure of the task is explicitly represented by a mental model. Of course these levels are not completely independent and tasks consist of actions from all three levels. But we propose emergency tasks can be classified according to the most central task requirements into two categories based on Rasmussen's taxonomy: *Rule-based* and *knowledge-based* tasks.

Effective performance in rule-based tasks is mostly based on rule-based behaviour and learned processes. During the task, specific signs (e.g. abnormal peaks in heart rate, drop in oxygen saturation) trigger stored rules (e.g. resuscitation algorithm or bag ventilation). The main requirement during these medical tasks is the correct execution of algorithm-based or prelearned manual actions. These tasks are characterized by low degrees of freedom (e.g. if ventricular fibrillation → then CPR, there is no other option). *Knowledge-based* tasks consist of processes on a higher cognitive level including identification of "symbols" that must be interpreted (Rasmussen, 1983). It is not just a signal that triggers a specific pattern of actions. Information has to be collected and interpreted, mental models about the condition of the patient have to be built, then a diagnosis has to be established and decisions about the treatment have to be made. In *knowledge-based* tasks, teams have higher degrees of freedom and there may be more than one correct way to treat the patient.

Coordination behaviour during rule- and knowledge based emergency tasks

A coordination behaviour investigated in the literature is task distribution. Task distribution includes all behaviours that assign tasks to team members (e.g. giving orders, delegating). The concept of task management that includes task distribution has been positively linked with performance in various studies (Burtscher et al., 2010; Manser et al., 2008). St. Pierre, Hofinger, Buerschaper & Simon (2011) state that one main task a leader has to perform during resuscitation is distributing tasks according to team members' individual skills and knowledge. Other studies linked task distribution by the team leader with performance in resuscitation tasks (Fernandez Castelao et al., 2011; Tschan, Semmer, Hunziker, & Marsch, 2011) but not in diagnostic tasks (Tschan et al., 2011). If tasks are not clearly distributed team members require high levels of situation awareness (Schulz, Endsley, Kochs, Gelb, & Wagner, 2013) and need to anticipate what tasks have to be done next and then take on the respective task. Because emergency scenarios are infrequent and often perceived as stressful team members do not have the cognitive resources necessary for anticipating required actions. Thus explicit task distribution is necessary. If tasks are effectively distributed the whole treatment will be faster and in emergency medicine performance is time critical. Based on these considerations we formulated the following hypothesis:

H1: Task distribution is positively related to clinical performance in emergency scenarios.

The literature showed that task distribution is especially effective in resuscitation scenarios (Fernandez Castelao et al. 2013; Tschan et al.

2011). A resuscitation scenario is based on predefined algorithms and is a typical rule-based task. In knowledge-based tasks task distribution should be less effective because the main focus of such tasks is not the execution of actions but the gathering and integration of information. Thus we formulate the following hypothesis:

H2: Task distribution is more effective in rule-based emergency tasks than in knowledge-based emergency tasks.

An other coordination behaviour mentioned in the literature and especially prominent in teamwork trainings is acknowledgement. We define acknow-ledgement as a confirmation if something has been understood. This includes the expressions “yes”, “ok”, “no” or the read-back of the previous statement as a confirmation (e.g. “yes, I will prepare the adrenalin” after a request for preparing adrenalin). Team members can acknowledge that they understood a) specific information or b) a specific task that they will execute. For the one who distributes a task, acknowledgement is a valuable feedback. A similar concept called “closed-loop communication” has been part of medical communication trainings (Brindley, 2010; McGinley & Pearse, 2007) and is widely used in aviation (Gladwell, 2008). The basic idea is that the receiver always has to acknowledge the understanding of the initial message (i.e. confirm or repeat the message). Although the closed-loop communication concept is widely used in medical teamwork trainings only little scientific evidence exists about the effectiveness of this concept (Härgestam, Lindkvist, Brulin, Jacobsson, & Hultin, 2013; Siassakos et al., 2011). Based on these considerations we propose the following hypothesis:

H3: Acknowledgement is positively linked with clinical performance in emergency scenarios.

Acknowledgement, as a reply to an order or a delegation, confirms that someone accepts responsibility for executing a task. If there is no acknowledgement it remains unclear if a specific task will be executed increasing the probability to miss an important action. Because rule-based tasks are more action oriented acknowledgement in rule-based tasks should result in more effective coordination. In knowledge-based tasks this relationship should be less string because these scenarios are less time critical and there is more time to understand, interpret and anticipate actions and information. Thus we formulated the following hypothesis:

H4: Acknowledgement is more effective in rule-based tasks than in knowledge-based tasks

The third coordination behaviour we investigate in the present study is providing information without request (PIWR). PIWR means all information that is given without someone explicitly asking for it. That may be information about the patient someone is noticing (e.g. "saturation is on 80%") or commenting owns action (e.g. "ok, I will now try to intubate the patient"). An overlapping concept that has been investigated already is talking to the room (Waller & Uitdewilligen, 2008). Talking to the room means unidirectional sharing of information to the room (Artman & Wærn, 1999). Tschan et al. (2009) found that talking to the room invites team members to focus on an issue and thus fosters a process of collaborative problem solving. This effect was found in diagnostic tasks but not in resuscitation tasks. Explicit sharing of unsolicited information with the rest of the team is important because it increases the possibility that the whole team has all necessary information to come up with the correct diagnosis. PIWR can help teams to build a shared mental model about the patient's condition (Burtscher & Manser, 2012) and may thus specifically increase team performance in knowledge-based scenarios where effective information processing should be a performance critical factor. We postulate that PIWR in rule-based scenarios is not linked with performance because the actions are driven by signals that trigger patterns of actions (i.e. treatment protocols). These signals are clearly linked with the action and no other information is needed to execute the correct treatment. Therefore we formulate the following hypothesis:

H5: PIWR is critical for high clinical performance only in knowledge-based emergency tasks but not in rule-based emergency tasks.

METHOD

This study was approved by the ethics committee of the Department of Psychology, University of Fribourg (Ref-Nr 2012_003R1). We obtained written informed consent from all participants that took part in the study.

Sample and procedure

We investigated 68 paediatric intensive care teams from seven different German hospitals. The study was part of in-house simulation trainings provided to the intensive care units (ICU) by PAEDSIM e.V. (www.paedsim.org) and took place between January 2012 and November 2013. Trainings were carried out in the real ICU environment using the local equipment and infrastructure. Per training session a group of 10-15 ICU staff was trained in three of four possible emergency scenarios. Out of this group four to eight people were chosen to deal with the patient always in an other constelation for each of the three scenarios. The teams were assembled according to the usual working conditions in the hospitals and consisted of one senior physician, one to three residents and one to three nurses. After every training scenario the trainer (E.H. or F. H.) debriefed the team. The debriefings as well as the trainings were based on self-reflection and medical skills so we do not expect a learning effect in terms of team coordination variables. Further there was no difference in performance related to the position of the training ($F(2,65) = 2.08$, $p = .13$)

A high-fidelity simulation mannequin of a 6 months old boy with highly realistic features was used to simulate the scenarios. The mannequin is able to show physical signs such as breathing and chest wall motion. Hearth rhythm is displayed after connecting the EKG, pulse can be palpated at several places and hearth and lung sound can be heard.

Task

Four different paediatric emergency scenarios were trained: Septic shock (severe blood infection; $N = 24$), bronchiolitis (illness of respiratory tract caused by infection of bronchioles; $N = 12$), near drowned child ($N = 13$) and pulseless ventricular tachycardia (PVT; live-threatening cardiac emergency; $N = 19$). All four scenarios are severe conditions and, without the appropriate treatment, will lead to death of the patient. All scenarios begin with a quick information exchange with an actor (mother or nurse handover) and then based on the condition of the patient the team should immediately start the treatment. The training scenarios last for 10-20 minutes depending on how fast the team achieves the treatment goals.

The classification of the four scenarios into rule- or knowledge-based tasks was done through a task analysis based on other studies (Schmutz & Manser, 2014). The drown child and PVT scenario were classified as rule-base scenarios. Both scenarios have clear triggers followed by strict treatment algorithms. In the PVT scenario a specific arrhythmia is visible on the heart monitor indicating the need for resuscitation following established resuscitation guidelines. In the drowned child scenario the child's oxygen saturation level drops below 80% (also visible on the monitor) indicating a need to use bag mask ventilation to re-establish the oxygen saturation of the patient and initiate the intubation procedure. Septic shock and bronchiolitis are both scenarios that are difficult to diagnose and are considered knowledge-based tasks based on the results of our task analysis. Several cues indicate a septic shock: lethargy, unresponsiveness, fever and dots on the skin. These symptoms need to be considered together with the fact that the child already has several hours of fever and vomiting. The patient in the bronchiolitis scenario is presented with increasing respiratory distress, is lethargic and has breathing noises that need to be detected. Further the fact that the child is a premature baby and has three days of rhinorrhoea, cough and progressive difficulty feeding need to be considered for correct diagnosis.

Measures

Coordination behaviour. Task distribution, acknowledgement and PIWR were coded with the CoMeT-E (Coordination System for Medical Teams - Emergency). The CoMeT-E is an adapted version of an established system originally developed for observing coordination in anaesthesia teams (Manser et al., 2008). It consists of 22 codes grouped into 9 main categories of which we considered task distribution, acknowledgement and PIWR for further analysis. Video coding was done by four trained organizational psychologist using the software Interact (Mangold International GmbH, Arnsdorf). We assessed the duration of every code and exported the amount of coordination (i.e. task distribution, acknowledgement, PIWR) in relation to the duration of the whole task. 15% of the videos were double coded by two raters to test interrater reliability. Cohens Kappa ranged from substantial to almost perfect (κ = 0.74 - 0.90).

Clinical performance. Clinical performance was assessed with a structured evaluation checklist. Because we wanted to assess scenario specific clinical performance we developed four individual checklists for each scenario according to an integrative approach for the development of clinical evaluation checklists (Schmutz, Eppich, Hoffmann, Heimberg, & Manser, 2014). Experienced clinicians rated all videos using the checklist. 15% of all videos were double rated. Interrater reliability was substantial to almost perfect (κ = 0.68 - 0.93). The evaluation checklists consisted of a similar structure (ABCD-assessment) but they differ in some scenario specific items resulting in different amount of checklist items. Therefore to make the performance measures comparable between the four scenarios we calculated the percentage of points in relation to the maximum amount of points possible.

Control variables. We included the following control variables in our analysis:

Leader experience: Previous research suggests that the leader may have an effect on clinical performance (Cooper & Wakelam, 1999). Experienced leaders have greater clinical knowledge and experience and thus might influence the clinical performance of the team. Mean paediatrics specific experience level of team leaders' in years was 12.4 years (*SD* = 8.2).

Team size: The size of the team ranged from four to eight clinicians which potentially could influence performance as well as coordination behaviour. Mean team size was 5.8 (*SD* = 1.0)

Duration of the scenario: Controlling for the duration of the scenario is important for two reasons. First out of educational reasons the trainers gave lower performing teams more time to solve the task instead of just stopping the scenario or even letting the patient die. Therefore the evaluation checklist includes time critical items so a delay in treatment will result in a lower performance score. Second, we expect the scenarios varying in duration with rule-based tasks generally requiring less time than knowledge-based tasks.

RESULTS

Table 1 provides means and standard deviations for the three coordination behaviours, clinical performance and the three control variables for rule- and skill-based tasks. Task distribution and acknowledgement occurred equally in both type of tasks. PIWR was observed significantly more in rule-based tasks than in knowledge-based tasks. Also the average team size was significantly higher in rule-based tasks. The duration for knowledge-based tasks was significantly longer than for rule-based tasks. All other variables were not significantly different in rule-based and knowledge-based tasks. Table 2 shows the correlation coefficients for all study variables.

We conducted hierarchical regression analysis. The predictor variables were centred around their grand mean to facilitate the interpretation of main effects in models containing interaction terms (Aiken & West, 1991). We calculated three separate regression models for task distribution, acknowledgement, and PIWR.

Table 3 provides the results of the three hierarchical regression analyses for the three coordination behaviours. The predictors were entered into the regression in the following three steps: (1) control variables: team size, scenario duration, leader experience; (2) task type and one of the three coordination behaviours (i.e., task distribution, acknowledgement or PIWR); (3) the interaction of task type and the respective coordination behaviour.

Table 1

Means and Standard Deviations Among All Study Variables for Rule- and Skill-Based Tasks

	Rule-based task (*N* = 32)		Knowledge-based task (*N* = 36)		Independent t-test		
	M	*SD*	*M*	*SD*	*t*	*df*	*p*
Task distribution[1]	23.46	5.19	23.02	6.15	0.32	66	.75
Acknowledgement[1]	8.83	3.00	9.84	4.38	-1.10	66	.28
PIWR[1]	34.88	13.71	17.03	6.76	6.92	66	<.01
Clinical performance[2]	41.34	10.86	42.97	11.38	-0.60	66	.55
Team size	6.31	0.97	5.42	0.91	3.95	66	<.01
Scenario duration[3]	697	179	906	213	-4.35	66	<.01
Leader experience[4]	13.06	8.14	11.83	8.40	0.61	66	.54

Note. PIWR = Provide information without request. [1]Numbers indicate percentage of time spent on coordination behavior in relation to the whole task; [2]Percentage of maximum clinical performance score; [3]Duration in seconds; [4]Leader experience specific to paediatrics in years.

TABLE 2

Correlations Among the Study Variables

Variables	1	2	3	4	5	6	7
Outcome variable							
1. Clinical performance							
Control variables							
2. Team size	-.16						
3. Scenario duration	-.35**	-.11					
4. Leader experience	.20*	.16	-.41**				
Predictor variables							
5. Task type[a]	.07	-.44**	.47**	-.08			
6. Task distribution	.16	.15	.09	.12	-.04		
7. Acknowledgement	.25*	-.15	.16	-.13	.14	.43**	
8. Provide information without request	-.05	.14	-.35**	-.11	-.65**	-.60	-.03

N = 68. [a] 0 = rule based scenario, 1 = knowledge based scenario. * $p < .05$, ** $p < .01$.

TABLE 3
Results of Hierarchical Regression Analyses for Predicting Teams Clinical Performance

	Model 1: [b]Task Distribution		Model 2: [b]Acknowledgement		Model 3: [b]PIWR	
	SE	β	*SE*	β	*SE*	β
Step 1: Control variables						
Constant	1.26		1.26		1.26	
Team size	1.25	-.21†	1.25	-.21†	1.25	-.21†
Scenario duration	.01	-.33*	.01	-.33*	.01	-.33*
Leader experience	.17	.10	.17	.10	.17	.10
ΔR^2	.17		.17		.17	
ΔF	4.21**		4.21**		4.21**	
Dfs	(3, 64)		(3, 64)		(3, 64)	
Step 2: Main effect						
Constant	2.06		2.01		2.49	
Team size	1.36	-.13	1.33	-.07	1.44	-.12
Scenario duration	.01	-.51**	.01	-.49**	.01	-.45**
Leader experience	.17	.00	.17	.07	.17	.05
Scenario type[a]	3.14	.26†	3.07	.24†	4.08	.19
Coordination[b]	.22	.24*	.32	.30**	.12	-.08
ΔR^2	.25		.29		.20	
ΔF	4.23**		4.93**		3.17*	
Dfs	(5, 62)		(5, 62)		(5, 62)	
Step 3: Moderation						
Constant	2.10		1.94		2.54	
Team size	1.37	-.13	1.35	.02	1.48	-.14
Scenario duration	.01	-.52**	.01	-.51**	1.48	-.45**
Leader experience	.17	-.00	.16	.01	.18	.04
Scenario type	3.22	.27†	2.97	.27†	4.50	-.15
Coordination[b]	.37	.29	.59	.72**	.14	-.03
Scenario type × Coordination[b]	.46	-.06	.71	-.49*	.31	-.10
ΔR^2	.27		.35		.21	
ΔF	3.49**		5.41**		2.66*	
Dfs	(6, 61)		(6, 61)		(6, 61)	

Note. [a] 0 = rule based-task, 1 = knowledge-based task; [b] coordination = 3 separate models for each coordination behaviour “task distribution”, “acknowledgement” or “PIWR” as predictor; † $p < .10$, * $p < 0.05$, ** $p < 0.01$.

Task distribution. As shown in step 2 in model 1 task distribution was positively related with clinical performance, (β = .24, p = .04). However, step 3 indicates no moderating effect of task type in this relationship (β = -.06, p = .74). Thus, hypothesis 1 is supported, but hypothesis 2 is not supported.

Acknowledgement. For model 2, step 2 indicates a positive effect of acknowledgement on task type (β = .30, p = .009). However, this main effect was moderated by task type in step 3 (β = -.49, p = .019). For a more specific test of hypothesis 4 we conducted a simple slope analysis according to Aiken and West (Aiken & West, 1991). The results indicate that acknowledgement was only positively related with performance in rule-based tasks (B = 2.09, SD = .59, p < .01) but not in knowledge-based tasks (B = .37, SE = .37, p = .32; see Figure 1). Therefore, hypothesis 3 and 4 were supported.

Provide information without request. In Model 3, there was neither a main effect of PIWR on clinical performance nor a moderation effect of task type (β = -.10, p = .59). Therefore, hypothesis 5 is not supported.

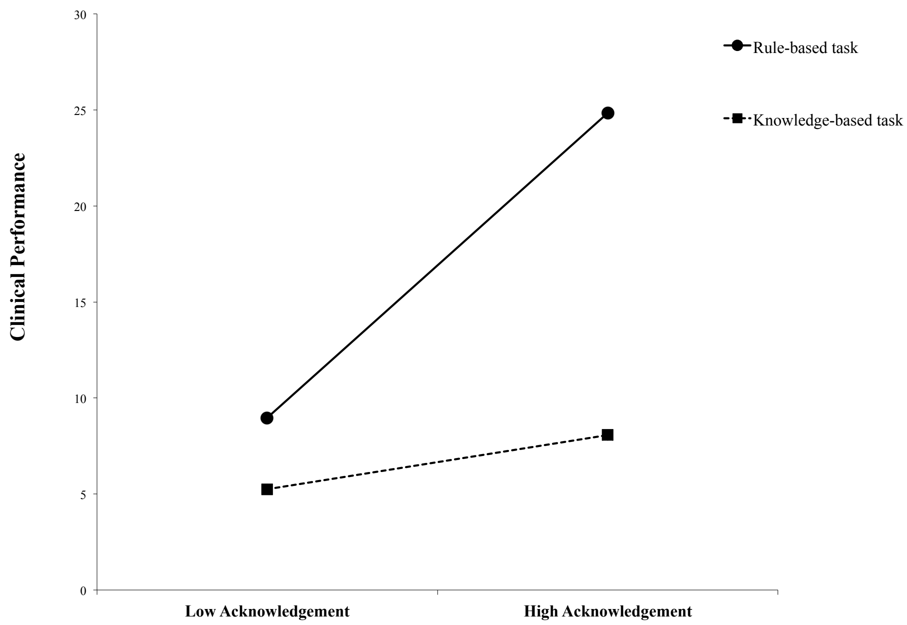

Figure 1. Interaction between acknowledgement and clinical performance for rule-based and knowledge-based tasks

DISCUSSION

The aim of the present study was to identify effective coordination behaviour of medical emergency teams that is adaptive to the task specific requirements. To do so we established two types of emergency tasks (i.e. rule-based and knowledge-based tasks) according to Rasmussen's taxonomy of human performance (Rasmussen, 1983). We were able to link two of three coordination behaviours commonly investigated in the literature with clinical performance of which acknowledgement was moderated by the task type.

Task distribution supported effective clinical performance for both task types and was not as expected more important in rule-based tasks. It thus seems that task distribution is a coordination behaviour that is effective for different task types. This does not necessarily mean that task distribution is a behaviour increasing the possibility of the team identifying task requirements such as planning or overall communication (Tschan et al., 2000). It rather seems that task distribution is important in response to emerging task requirements in different emergency scenarios. A team can only function if all members know what needs to be done, when, how and who is doing it (Tschan, 2002). Task distribution ensures that the roles and responsibilities in the team are made explicit. Although we defined two types of emergency tasks they can both be seen as non-routine tasks where task distribution[5] (and explicit coordination in general) is more effective than implicit coordination (Zala-Mezo et al., 2009). Tasks may either be distributed hierarchically from the senior physician to the resident or nurses or laterally from one nurse to another (e.g. "you prepare the heparin and I will call for help"). Emergency tasks usually take place under high time pressure where different tasks need to be done by different team members at the same time. Therefore to synchronize all team actions and to take into account individual's competencies task distribution is needed for a smooth functioning of the team.

After discussions with trainees we learned that some of them were not aware of what medical illness the patient was present until the end of the training scenario. This indicates that there is

[5] According to our definition task distribution is clearly an explicit coordination mechanism.

not necessarily a shared mental model present that would be a prerequisite for anticipating tasks and making explicit task distribution unnecessary (Burtscher & Manser, 2012). Overall roles and responsibilities in the team are predefined by professional backgrounds (e.g. physician, nurse) and all team members know, for example, that the nurses prepare the medication and physicians are responsible to assess the patient. But still nurses have to rely on physicians' orders concerning which medication and what dose exactly they have to prepare. The subtasks at hand are not implicitly distributed as in routine situations such as regular paediatric ward care where nurses work independently for much longer periods and only coordinate with physicians during rounds.

To our knowledge this is one of the first studies to empirically link acknowledgement with clinical performance in healthcare. Our results indicate that acknowledgement is important in emergency scenarios in general but that this effect is more pronounced for rule-based tasks. Acknowledgement is defined as a confirmation that a specific piece of information or a task assignment has been understood and, in case of a task, will be executed. Acknowledgement closes the communication loop and can prevent communication breakdowns and misunderstandings (e.g. one team member assuming that the other will give a drug while the other person did not hear that order). As can be seen in Table 2 acknowledgement is positively related to task distribution. This suggests that acknowledgement might frequently be a response to task distribution in the sense of the closed loop communication model (Härgestam et al., 2013). Such a confirmation might be especially relevant for rule-based tasks because the fast and accurate execution of critical actions is the main goal in these situations. Future studies should investigate this sequential link in more detail.

Our hypothesis that PIWR was related to clinical performance in knowledge-based tasks but not in rule-based tasks was not supported. It is notable that in rule-based tasks there is a higher amount of PIWR than in knowledge-based tasks. This can partly be explained by team members explicitly commenting on their own actions during manual tasks (e.g. commenting on actions during intubation, counting out loud during CPR). The intention of this behaviour is to get all team members informed so they can anticipate the next step (e.g. bag ventilation after 15 chest compressions). We propose three possible reasons why PIWR was not related to effective performance in our emergency tasks.

First, we defined PIWR as all undirected talk and sharing of information. In doing so a differentiation whether this was related to information (e.g. heart rate is 150) or tasks (e.g. "I just gave the medication") is lacking. Kolbe et al. (2010) did investigate the similar concept of talking to the room and found different effects for information related and task related talking to the room. One could expect that for the knowledge-based tasks information related PIWR is related to a faster interpretation of patient information and thus to a more accurate treatment. Task related PIWR might in this case distract the team from the main task to gather and interpret all necessary information and thus hinder clinical performance.

Second, we employed a purely descriptive observation system for coordi-nation behaviour. There is no evaluative component like in other observation systems such as NOTECHS (Mishra, Catchpole, & McCulloch, 2009). Thus, we did not assess if the information provided was needed and valuable for the team. If a team member provides information that is already present in the team or deemed to be irrelevant at this point in time the team might not benefit form PIWR but rather be distracted from the main task. Future studies should investigate PIWR taking into account the broader context and considering the goals and intentions of other team members as, for example, in the concept of heedful interrelating (Weick & Roberts, 1993) that assumes that providing specific information can be heedful or not depending on the situation and the information needs of team members.

Third, PIWR is not limited by the situation itself in the same way as task distribution or acknowledgement is. Acknowledgement is a reaction to an order, delegation or information. A team member cannot acknowledge without a preceding coordination behaviour that implies acknowledgement. Thus, team members cannot show more acknowledgement than necessary. Similarly, task distribution is limited to the tasks that need to be performed in a given situation. If all tasks are distributed appropriately, thus supporting effective performance, team members are unlikely to engage in task distribution. In contrast, PIWR is not naturally limited by the task and may only be supporting effective performance up to a certain degree after which it may contribute to communication overload and harm performance. Thus, a non-linear relationship of PIWR and clinical performance might be more accurate.

Exploratory analysis of our data even indicated a tendency towards a negative relationship of PIWR and clinical performance; in particular in teams with an experienced leader. Future studies will have to show whether an experienced leader is in fact less likely to benefit form PIWR and if there is a tipping point where too much information can even be detrimental to performance.

In addition our results support the distinction into rule- and knowledge-based tasks. Knowledge-based tasks take longer which is related with the fact that gathering and integrating information in the knowledge-based tasks takes more time whereas the rule-based tasks can be seen as shorter and more intense tasks. In rule-based tasks there are usually more team members involved than in knowledge-based tasks. An emergency team in a hospital always consists of a core team but always has the possibility to ask for help of additional staff members. To be as realistic as possible the teams had the opportunity to call two to four more nurses for help, as they would do in real life if necessary. In rule-based tasks it makes sense to call for help to have more resources to manage the task. However, in knowledge-based tasks it is not necessary to call for more team members because it is mainly the senior physician and the resident who collect the information and come up with a diagnosis, so there is only limited need for calling more nursing staff. Further it is known that standardized procedures support or can even substitute coordination (Van de Ven, Delbecq, & Koenig Jr, 1976). Thus, it might be easier to coordinate a big team in a rule-based scenario where coordination is driven by algorithms and just has to be updated among team members, whereas the coordination effort in a knowledge-based task would exceed the benefits in a bigger team.

Further we would like to emphasize that our results concerning the effectiveness of task distribution and acknowledgement are independent of team leader experience. One might assume that in teams with different experience levels of the leader different coordination behaviours are effective. In our study this was not the case, at least for task distribution and acknowledgement.

Thus, these two coordination behaviours should be used and trained independent of team leader's experience level.

Limitations and strengths of the study

First, in our coding procedure we were not able to link coordination behaviours with a person. Identification of the speaker was difficult because there was a considerable amount of overlapping coordination making it difficult to distinguish speakers and lip movements were not always visible because team members were mostly bent over the patient. Linking an observation code with a person might be easier in live observation because a) it would be possible to move if something is not visible and b) communication is more audible than on audio recordings. However, given that our data was collected during in-situ trainings of actual paediatric emergency teams across Germany, coding of video recordings was more feasible.

Second, the emergency tasks were performed with a paediatric simulator. Although the manikin is highly realistic it cannot be ruled out that teams would behave differently in real life (e.g. due to a higher stress level) and under no observation. Nevertheless, one big advantage of using a simulator is the standardization of the specific tasks because in real life it is almost impossible to observe teams treating patients with exactly the same conditions.

Third, the teams participating in this study were very heterogeneous in terms of size and composition. This might be seen as a limitation, but we regard it also as strength because it reflects the real working conditions of medical emergency teams. There is no predefined amount of team members for emergency teams. This depends on time (e.g. day vs. night shift), size of the whole ward team and the workload of other colleagues. An emergency team is always assembled ad hoc in case of an emergency.

Along these lines the major strength of this study is the investigation of fully qualified clinicians as they work together in their everyday work and in their regular environment with their own equipment. Many studies investigate healthcare teams consisting of medical students or only homogenous professional groups (e.g. only physicians representing an interprofessional treatment team) limiting the ecological validity of results.

Compared to other studies in the field we investigated a rather large sample of teams and the fact that we did not compare only two specific tasks but four different emergencies increases the generalizability of our result. Furthermore to our knowledge this is the first study that could show a moderation of task type in relation with team coordination and performance.

Practical implications

Our results have important practical implications for the work and training of emergency teams and thus for patient safety. We were able to show, as well as other studies before, that not only the technical skills of healthcare teams are important for clinical performance but also non-technical skills such as coordination. Our measure of clinical performance of a team is based on international medical guidelines and thus can make the difference between life and death of a patient in an emergency situation.

Task distribution and acknowledgement should be integrated into teamwork trainings and thus enhance clinical team performance and patient safety. Because emergency situations are highly stressful situations it is not enough to just train these coordination behaviours once. Rather it is necessary to train them regularly and integrate them into everyday work practices.

Further our results showed that PIWR is not necessarily an effective coordination strategy. Healthcare workers should be aware that providing unsolicited information to the team is not always supporting effective team performance. Teams should not simply use PIWR but consider this coordination behaviour in the context of the task and goals of the team and only use it if the team also may benefit from the delivered information (e.g. the specific information is clearly missing in the team and important for goal attainment or the overall communication level is not yet too high).

Healthcare workers and trainers should be aware that there are different task types that place different coordination requirements on the team. Our results indicate that acknowledgement is especially important in rule-based tasks so this behaviour should be an important element in resuscitation trainings and other rule-based tasks.

In sum we provide valuable results for three concrete coordination behaviours that can and should be trained in emergency team trainings. With a rather large sample compared to other studies in the field our results are a strong indicator for the importance of coordination in healthcare teams and we hope that these results will make a contribution to improve team coordination and thus make healthcare safer.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems: Formation, coordination, development, and adaptation*. Thousand Oaks, CA: Sage.

Artman, H., & Wærn, Y. (1999). Distributed cognition in an emergency co-ordination center. *Cognition, Technology & Work, 1*, 237-246.

Boos, M., Kolbe, M., & Strack, M. (2010). An inclusive model of group coordination. In M. Boos, M. Kolbe, P. M. Kappeler, & T. Ellwart (Eds.), *Coordination in human and primate groups* (pp. 11-35). Berlin: Springer.

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement: Theory, methods, and applications. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team Performance Assessment and Measurement* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum.

Brindley, P. G. (2010). Communication in crisis: The importance of 'Verbal Dexterity'. In J.-L. Vincent (Ed.), *Intensive Care Medicine* (pp. 581-589). Berlin: Springer.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *The Journal of Applied Psychology, 91*, 1189-1207.

Burtscher, M. J., & Manser, T. (2012). Team mental models and their potential to improve teamwork and safety: A review and implications for future research in healthcare. *Safety Science, 50,* 1344-1354.

Burtscher, M. J., Manser, T., Kolbe, M., Grote, G., Grande, B., Spahn, D. R., et al. (2011). Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. *British Journal of Anaesthesia, 106*, 1-6.

Burtscher, M. J., Wacker, J., Grote, G., & Manser, T. (2010). Managing nonroutine events in anesthesia: The role of adaptive coordination. *Human Factors, 52*, 282-294.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. *Team Effectiveness and Decision Making in Organizations, 333*, 380.

Cooper, S., & Wakelam, A. (1999). Leadership of resuscitation teams: "Lighthouse Leadership". *Resuscitation, 42*, 27-45.

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.

Entin, E. E., & Serfaty, D. (1999). Adaptive Team Coordination. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41*, 312-325.

Fernandez Castelao, E., Russo, S. G., Cremer, S., Strack, M., Kaminski, L., Eich, C., et al. (2011). Positive impact of crisis resource management training on no-flow time and team member verbalisations during simulated cardiopulmonary resuscitation: A randomised controlled trial. *Resuscitation, 82*, 1338-1343.

Fernandez Castelao, E., Russo, S. G., Riethmüller, M., & Boos, M. (2013). Effects of team coordination during cardiopulmonary resuscitation: A systematic review of the literature. *Journal of Critical Care, 28*, 504-521.

Gladwell, M. (2008). The ethnic theory of plane crashes *Outliers* (pp. 177-223). New York, NY: Little, Brown and Company.

Grote, G., Kolbe, M., Zala-Mezo, E., Bienefeld-Seall, N., & Kunzle, B. (2010). Adaptive coordination and heedfulness make better cockpit crews. *Ergonomics, 53*, 211-228.

Härgestam, M., Lindkvist, M., Brulin, C., Jacobsson, M., & Hultin, M. (2013). Communication in interdisciplinary teams: exploring closed-loop communication during in situ trauma team training. *British Medical Journal Open, 3*, e003525.

Hunziker, S., Bühlmann, C., Tschan, F., Balestra, G., Legeret, C., Schumacher, C., et al. (2010). Brief leadership instructions improve cardiopulmonary resuscitation in a high-fidelity simulation: A randomized controlled trial*. *Critical Care Medicine, 38*, 1086-1091.

Kolbe, M., Burtscher, M., Manser, T., Kunzle, B., & Grote, G. (2010). The role of coordination in preventing harm in healthcare groups: Research examples from anaesthesia and an integrated model of coordination for action teams in health care. In M. Boos, M. Kolbe, P. M. Kappeler, & T. Ellwart (Eds.), *Coordination in human and primate groups* (pp. 75-92). Berlin: Springer.

Kolbe, M., Künzle, B., Zala-Mezö, E., Burtscher, M. J., Wacker, J., Spahn, D. R., et al. (2010). *The functions of team monitoring and 'talking to the room'for performance in anesthesia teams.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Manser, T., Harrison, T. K., Gaba, D., M., & Howard, S. K. (2009). Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesthesia & Analgesia, 108*, 1606-1615.

Manser, T., Howard, S. K., & Gaba, D. M. (2008). Adaptive coordination in cardiac anaesthesia: A study of situational changes in coordination patterns using a new observation system. *Ergonomics, 51*, 1153-1178.

McGinley, A., & Pearse, R. M. (2007). Medical emergency teams: Implementation and outcome measurement. *Critical Care, 11*, 307.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

Mishra, A., Catchpole, K., & McCulloch, P. (2009). The Oxford NOTECHS System: Reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality and Safety in Health Care, 18*, 104-108.

Parker, S. E. H., Schmutz, J., & Manser, T. (under review). Coordination in healthcare action teams: Utilizing expert understanding of task and team performance requirements.

Rall, M., Gaba, D. M., Howard, S. K., & Dieckmann, P. (2005). Human performance and patient safety. *Miller's anesthesia. Elsevier, Philadelphia*, 3021-3072.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, 13*, 257-266.

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2009). Developing a team performance framework for the intensive care unit. *Critical Care Medicine, 37*, 1787-1793.

Rico, R., Sanchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge-based approach. *The Academy of Management Review, 33*, 163-184.

Riethmüller, M., Fernandez Castelao, E., Eberhardt, I., Timmermann, A., & Boos, M. (2012). Adaptive coordination development in student anaesthesia teams: A longitudinal study. *Ergonomics, 55*, 55-68.

Schmutz, J., Eppich, W. J., Hoffmann, F., Heimberg, E., & Manser, T. (2014). Five steps to develop checklists for evaluating clinical performance: An integrative approach. *Academic Medicine, 89*.

Schmutz, J., & Manser, T. (2013). Do team processes really have an effect on clinical performance? A systematic literature review. *British Journal of Anaesthesia, 110*, 529-544.

Schmutz, J., & Manser, T. (2014, March). *Using hierarchical task analysis to identify different coordination requirements in a pediatric emergency scenario.* Paper presented at the International Symposium on Human Factors and Ergonomics in Health Care, Chicago, Ill.

Schulz, C. M., Endsley, M. R., Kochs, E. F., Gelb, A. W., & Wagner, K. J. (2013). Situation awareness in anesthesia: concept and research. *Anesthesiology, 118*, 729-742.

Siassakos, D., Fox, R., Crofts, J. F., Hunt, L. P., Winter, C., & Draycott, T. J. (2011). The management of a simulated emergency: Better teamwork, better performance. *Resuscitation, 82*, 203-206.

St. Pierre, M., Hofinger, G., Buerschaper, C., & Simon, R. (2011). *Crisis management in acute care settings - Human factors and team psychology in a high stakes environment*. Berlin: Springer.

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and Effectiveness. *American psychologist, 45*, 120.

Tschan, F. (2002). Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research, 33*, 615-643.

Tschan, F., Semmer, N., Vetterli, M., Gurtner, A., Hunziker, S., & Marsch, S. U. (2011). Developing Observational Categories for Group Process Research Based on Task and Coordination Requirement Analysis: Examples from Research on Medical Emergency-Driven Teams. In M. Boos, M. Kolbe, P. Kappeler, & T. Ellwart (Eds.), *Coordination in human and primate groups* (pp. 93-115). Berlin: Springer.

Tschan, F., Semmer, N. K., Gurtner, A., Bizzari, L., Spychiger, M., Breuer, M., et al. (2009). Explicit Reasoning, Confirmation Bias, and Illusory Transactive Memory: A Simulation Study of Group Medical Decision Making. *Small Group Research, 40*, 271-300.

Tschan, F., Semmer, N. K., Hunziker, S., & Marsch, S. (2011). Decisive action vs joint deliberation: Different medical tasks imply different coordination requirements. In Duffy (Ed.), *Advances in Human Factors and Ergonomics in Healthcare.* Boca Raton, FL: Taylor & Francis.

Tschan, F., Semmer, N. K., Nägele, C., & Gurtner, A. (2000). Task adaptive behavior and performance in groups. *Group Processes & Intergroup Relations, 3*, 367-386.

Van de Ven, A. H., Delbecq, A. L., & Koenig Jr, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, *41,* 322-338.

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.

Waller, M. J., & Uitdewilligen, S. (2008). Talking to the room. Collective sensemaking during crisis situations. In R. Roe, M. Waller, & S. Clegg (Eds.), *Time in organizations - Approaches and methods* (pp. 186-203). London: Routledge.

Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, *38,* 357-381.

Zala-Mezo, E., Wacker, J., Kunzle, B., Bruesch, M., & Grote, G. (2009). The influence of standardisation and task load on team coordination patterns during anaesthesia inductions. *Quality & Safety in Health Care, 18*, 127-130.

Thesis References

Anderson, N., Ones, D. S., Sinangil, H. K., & Viswesvaran, C. (2001). *Handbook of industrial, work and organizational psychology: Personnel psychology* (Vol. 1). London: Sage Publications Ltd.

Annett, J. (2003). Hierarchical task analysis. *Handbook of cognitive task design*, 17-35.

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems: Formation, coordination, development, and adaptation*. Thousand Oaks, CA: Sage.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193,* 31-35.

Bobbio, A., Bellan, M., & Manganelli, A. M. (2012). Empowering leadership, perceived organizational support, trust, and job burnout for nurses: A study in an Italian general hospital. *Health Care Management Review, 37*, 77–87.

Brannick, M. T., & Prince, C. (1997). An Overview of Team Performance Measurement: Theory, Methods, and Applications. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum.

Bratt, M. M., Broome, M., Kelber, S., & Lostocco, L. (2000). Influence of stress and nursing leadership on job satisfaction of pediatric intensive care unit nurses. *American Journal of Critical Care, 9*, 307–317.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *The Journal of Applied Psychology, 91*, 1189-1207.

Burtscher, M. J., Kolbe, M., Wacker, J., & Manser, T. (2011). Interactions of team mental models and monitoring behaviors predict team performance in simulated anesthesia inductions. *Journal of Experimental Psychology: Applied, 17*, 257.

Burtscher, M. J., Manser, T., Kolbe, M., Grote, G., Grande, B., Spahn, D. R., & Wacker, J. (2011). Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. *British Journal of Anaesthesia, 106*, 1-6.

Burtscher, M. J., Wacker, J., Grote, G., & Manser, T. (2010). Managing nonroutine events in anesthesia: The role of adaptive coordination. *Human Factors, 52*, 282-294.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Cascio, W. F. (1987). *Applied psychology in personnel management*: Prentice-Hall Englewood Cliffs, NJ.

Catchpole, K. R., Giddings, A. E., Wilkinson, M., Hirst, G., Dale, T., & de Leval, M. R. (2007). Improving patient safety by identifying latent failures in successful operations. *Surgery, 142*, 102-110.

Cooper, S., & Cade, J. (1997). Predicting survival, in-hospital cardiac arrests: Resuscitation survival variables and training effectiveness. *Resuscitation, 35*, 17-22.

Davenport, D. L., Henderson, W. G., Mosca, C. L., Khuri, S. F., & Mentzer Jr, R. M. (2007). Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *Journal of the American College of Surgeons, 205*, 778-784.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95*, 32.

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement* (pp. 19-43). Mahwah, NJ: Lawrence Erlbaum.

Diener, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. (1976). Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *Journal of Personality and Social Psychology, 33*, 178.

Donoghue, A., Nishisakia, A., Suttona, R., Halesc, R., & Boulet, J. R. (2010). Reliability and validity of a scoring instrument for clinical performance during Pediatric Advanced Life Support simulation scenarios. *Resuscitation, 81*, 331-336.

Donoghue, A., Ventre, K., Boulet, J., Brett-Fleegler, M., Nishisaki, A., Overly, F., & Cheng, A. (2011). Design, implementation, and psychometric analysis of a scoring instrument for simulated pediatric resuscitation: A report from the EXPRESS pediatric investigators. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, 6*, 71-77.

DuPaul, G. J., Rapport, M. D., & Perriello, L. M. (1991). Teacher ratings of academic skills: The development of the Academic Performance Rating Scale. *School Psychology Review, 20*, 284-300.

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.

Fernandez Castelao, E., Russo, S. G., Riethmüller, M., & Boos, M. (2013). Effects of team coordination during cardiopulmonary resuscitation: A systematic review of the literature. *Journal of Critical Care, 28*, 504-521.

Gell-Mann, M. (1995). What is complexity? Remarks on simplicity and complexity by the Nobel Prize-winning author of *The Quark and the Jaguar*. *Complexity, 1*, 16-19.

Gordon, T. J. (1994). *The Delphi Method*. Futures Research methodology.

Grote, G., Kolbe, M., Zala-Mezo, E., Bienefeld-Seall, N., & Kunzle, B. (2010). Adaptive coordination and heedfulness make better cockpit crews. *Ergonomics, 53*, 211-228.

Guzzo, R. A., & Shea, G. P. (1992). Group performance and intergroup relations in organizations. *Handbook of industrial and organizational psychology, 3*, 269-313.

Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice-Hall.

Härgestam, M., Lindkvist, M., Brulin, C., Jacobsson, M., & Hultin, M. (2013). Communication in interdisciplinary teams: Exploring closed-loop communication during in situ trauma team training. *BMJ open, 3*, e003525.

Holmboe, E. S., & Hawkins, R. E. (2008). *Practical guide to the evaluation of clinical competence*. Philadelphia, PA: Mosby/Elsevier.

Ilgen, D. R. (1999). Teams embedded in organizations: Some implications. *American Psychologist, 54*, 129.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology, 56*, 517-543.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.* Boston, MA: Houghton Mifflin Company.

Kozlowski, S. W., Gully, S. M., Nason, E. R., & Smith, E. M. (1999). Developing adaptive teams: A theory of compilation and performance across levels and time. In D. R. Ilgen & E. D. Pulakos (Eds.), *The Changing Nature of Performance* (pp. 240-292). San Francisco, CA: Jossey-Bass.

Landy, F. J., & Conte, J. M. (2009). *Work in the 21st century: An introduction to industrial and organizational psychology*. Hoboken, NJ: John Wiley & Sons.

Lehmann-Willenbrock, N., Lei, Z. K., & Kauffeld, S. (2012). Appreciating age diversity and German nurse well-being and commitment: Co-worker trust as the mediator. *Nursing & Health Sciences, 14*, 213–220.

Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology, 41*, 585-634.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Manser, T. (2009). Teamwork and patient safety in dynamic domains of healthcare: A review of the literature. *Acta Anaesthesiologica Scandinavica, 53*, 143-151.

Manser, T., Howard, S. K., & Gaba, D. M. (2008). Adaptive coordination in cardiac anaesthesia: A study of situational changes in coordination patterns using a new observation system. *Ergonomics, 51*, 1153-1178.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review, 26*, 356-376.

McCulloch, P., Mishra, A., Handa, A., Dale, T., Hirst, G., & Catchpole, K. (2009). The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Quality & Safety in Health Care, 18*, 109-115.

McGrath, J. E. (1964). *Social psychology: A brief introduction*. New York: Holt, Rinehart and Winston.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Englewood Cliffs, NJ: Prentice-Hall.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review, 4*, 95-105.

Nerdinger, F. W., Blickle, G., & Schaper, N. (2011). *Arbeits- und Organisationspsychologie*. Berlin: Springer.

Parker, S. E. H., Schmutz, J., & Manser, T. (under review). Coordination in healthcare action teams: Utilizing expert understanding of task and team performance requirements.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*: Clarendon Press Oxford.

Pronovost, P. J., & Freischlag, J. A. (2010). Improving Teamwork to Reduce Surgical Mortality. *Journal of the American Medical Association, 304*, 1721-1722.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics,13*, 257-266.

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2009). Developing a team performance framework for the intensive care unit. *Critical Care Medicine, 37*, 1787-1793.

Riethmüller, M., Fernandez Castelao, E., Eberhardt, I., Timmermann, A., & Boos, M. (2012). Adaptive coordination development in student anaesthesia teams: A longitudinal study. *Ergonomics, 55*, 55-68.

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Cambridge, MA: Harvard University Press.

Salas, E., Rosen, M. A., & King, H. (2007). Managing teams managing crises: Principles of teamwork to improve patient safety in the emergency room and beyond. *Theoretical Issues in Ergonomics Science, 8*, 381-394.

Schmutz, J., Eppich, W. J., Hoffmann, F., Heimberg, E., & Manser, T. (2014). Five steps to develop checklists for evaluating clinical performance: An integrative approach. *Academic Medicine, 89*.

Schmutz, J., Hoffmann, F., Heimberg, E., & Manser, T. (under review). Effective Team Coordination in Emergency Care: The Moderating Role of the Task.

Schmutz, J. & Manser, T. (2013). Do team processes really have an effect on team performance? A systematic literature review. *British Journal of Anaesthesia, 110,* 529-544.

Schulz, C. M., Endsley, M. R., Kochs, E. F., Gelb, A. W., & Wagner, K. J. (2013). Situation awareness in anesthesia: Concept and research. *Anesthesiology, 118*, 729-742.

Shaw, M. E. (1971). *Group dynamics: The psychology of small group behavior*. New York City, NY: McGraw-Hill.

Siassakos, D., Fox, R., Crofts, J. F., Hunt, L. P., Winter, C., & Draycott, T. J. (2011). The management of a simulated emergency: Better teamwork, better performance. *Resuscitation, 82*, 203-206.

Sonnentag, S., & Frese, M. (2002). Performance concepts and performance theory. In S. Sonnentag (Ed.), *Psychological management of individual performance* (pp. 1-25). West Sussex: John Wiley & Sons, LTD.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.

Stufflebeam, D. L. (2000). Guidelines for developing evaluation checklists: The checklists development checklist (CDC). Retrieved May 24, 2014, from http://www.wmich.edu/evalctr/archive_checklists/guidelines_cdc.pdf.

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist, 45*, 120.

Tschan, F., Semmer, N. K., Hunziker, S., & Marsch, S. (2011). Decisive action vs joint deliberation: Different medical tasks imply different coordination requirements. In Duffy (Ed.), *Advances in human factors and ergonomics in healthcare*. Boca Raton: Taylor & Francis.

Tschan, F., Semmer, N. K., Nägele, C., & Gurtner, A. (2000). Task adaptive behavior and performance in groups. *Group Processes & Intergroup Relations, 3*, 367-386.

Viswesvaran, C. (2001). Assessment of individual job performance: A review of the past century and a look ahead *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 110-126). London: Sage Publication Ltd.

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.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*, 1036-1039.

Ich erkläre ehrenwörtlich, dass ich meine Dissertation selbständig und ohne unzulässige fremde Hilfe verfasst habe und sie noch keiner anderen Fakultät vorgelegt habe.

Jan Schmutz,
Fribourg, 26.5.2014